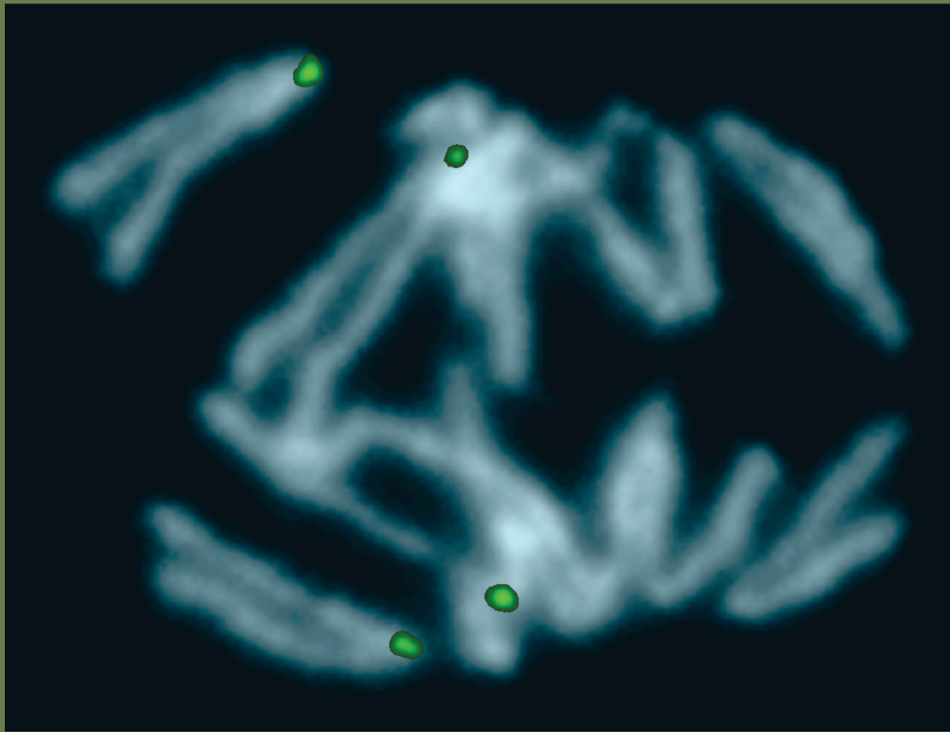


Análisis estructural del centrómero en *Drosophila melanogaster*



María Méndez Lago



Facultad de Ciencias
Departamento de Biología

Análisis estructural del centrómero en *Drosophila melanogaster*

Memoria presentada por
MARÍA MÉNDEZ LAGO
Madrid, 2008

Universidad Autónoma de Madrid
Facultad de Ciencias
Departamento de Biología

Análisis estructural del centrómero en *Drosophila melanogaster*

Tesis Doctoral presentada por
MARÍA MÉNDEZ LAGO
Para optar al grado de Doctor

Vº Bº Autora

Fdo: María Méndez Lago

Vº Bº Director de la Tesis Doctoral

Vº Bº Tutor

Dr. Alfredo Villasante Atienza
Centro de Biología Molecular Severo Ochoa
Universidad Autónoma de Madrid

Dr. Carlos Sentís Castaño
Departamento de Biología
Universidad Autónoma de Madrid

A Toñi

AGRADECIMIENTOS

Éste es, para mí, el apartado más importante que escribir en la tesis, pues si bien no refleja la importancia o relevancia del trabajo científico realizado, sí refleja la calidad humana de todas las personas que me habéis apoyado a recorrer este camino de una u otra forma, a veces sin ser conscientes, dándome fuerzas y energías. Por eso, no sería justo terminar esta tesis sin hacer mención a todos vosotros y daros mi más sincero agradecimiento por vuestra contribución.

En primer lugar quiero agradecer de corazón a Toni por haber confiado en mí y haberme propuesto realizar esta tesis doctoral. Pero, sobre todo, quiero agradecerle por hacer de mí “una de tus niñas”, por haberme enseñado muchas cosas importantes en la vida, como a disfrutar y a luchar. Me encantaría que hoy estuvieses aquí para ver el resultado de todos estos años; sé que estarías orgullosa. También quiero agradecer a Gracia, porque ella me ha seguido transmitiendo todo su cariño y su apoyo, y a través del suyo, el tuyo también. Gracias.

En segundo lugar, y no menos importante, quiero agradecer a Alfredo, por todo. Por su apoyo en todo momento, por su confianza infinita en mí (creo que, a veces, exagerada), por haberme escuchado y haberme entendido, por haberme enseñado y haberme dirigido en el trabajo de esta tesis, y por el cariño con que lo ha hecho, casi más como un padre que como un mero director de tesis.

Mil gracias también a Bea, por toda su ayuda, que no ha sido poca. Por todas las horas compartidas en el laboratorio, de trabajo y de risas y anécdotas. Por su calma, por sus conocimientos, por su amistad. Y también a Pascual, a Charo, a Mari Carmen y a Paula, por el apoyo recibido por todos ellos y el tiempo compartido en el laboratorio. A Esteban, por su apoyo y cariño, mostrados a su manera, pero bien recibidos. Y a Carmen, por esta nueva amistad y apoyo constante en la última fase de la tesis.

A Wacław Szybalski, por permitirme realizar una estancia en su laboratorio, por enseñarme muchas cosas sobre la vida científica y por compartir sus vivencias conmigo. A Jadwiga, por haberme enseñado tantas cosas, como científica y como “madre adoptiva”. Por su cariño y por sus reflexiones compartidas. Y a Zdenka, por su permanente sonrisa, y por enseñarme que la edad no importa, ni para la ciencia ni para la vida en general. A Kelly Winterberg y a Reznikoff, por facilitarme la transposasa y su protocolo. A Dan Loab, por prestarme un espacio en su laboratorio para los experimentos con radioactividad. A Dong Yun, por su compañía durante mi estancia y su amistad que perdura.

A Gary Karpen, y a todos los miembros de su laboratorio en Berkeley, por tenerme allí durante unos meses y por hacerme sentir como una más del laboratorio. Especialmente a Matt y a Patrick, por

enseñarme a obtener fibras de cromatina, y a Bárbara, por su amistad. A Susan Celniker y a Roger Hoskins, por hacerme un hueco también en su laboratorio y ayudarme con la secuenciación y análisis de los BACs. A Ken, por su simpatía especial.

A Siobhan Whitehead por apoyarme desde la distancia con los programas de ensamblaje de secuencias, por invitarme al Sanger Institute a resolver las cuestiones inabordables por teléfono. Y por su empeño y esfuerzo, junto con el de Alan Tracey, para secuenciar los BACs heterocromáticos.

A Francisco Mampaso, del hospital Ramón y Cajal, por cederme su termociclador para PRINS, y a Marta, por acogerme en su laboratorio. ¡Qué placer trabajar contigo, a pesar de tener que ver lo que les haces a los ratoncitos!

A Olga Demakova, del laboratorio de Igor Zhimulev, por hacer las “in situs” sobre cromosomas politénicos en la cepa mutda para Su(var)3-9 y SuUR.

A Bernardo López por instruirme sobre los distintos métodos interpoladores, y por enseñarme a utilizar el programa Matlab para realizar los “splines”.

A Carlos Sentís, por ser mi tutor de tesis y por sus consejos.

A la Fundación Ramón Areces y al CSIC por otorgarme sendas becas para la realización de la investigación predoctoral. Y a la Fundación Fulbright, y al Instituto Internacional en España, por concederme la bolsa de viaje que permitió mi estancia en la Universidad de Wisconsin.

Volviendo al plano personal, quiero agradecer con especial fuerza a Berta, mi madre, a Juan Luis, mi padre, a mi abuelo Antonio y a mis hermanas Mónica y Ana. Han seguido de cerca (o lejos) toda mi tesis y, sin entender ni palabra, han mostrado un gran interés día tras día por saber cómo avanzaba. Me han dado todo su apoyo en múltiples formas, me han cuidado y se han preocupado por mí. Sin su apoyo y su amor esto hubiese sido mucho más duro. Muchísimas gracias. Bueno, y gracias también al resto de la familia, a Nati, Margot, a todos los tíos y primos, que, en lo cotidiano, también habéis ayudado. Y a Elisa, que tanta ilusión y energía me ha dado en la última fase de la tesis, y ella, seguro, sin saberlo.

Por supuesto, debo millones de gracias a “las chicas SIdI”: Conchi, Tais y Charo. No sé ni cómo deciros todo lo que os debo, todo lo que en mi tesis va de vosotras (a parte de las secuencias, claro está). Sé que no hace falta, porque esta tesis es casi tan vuestra como mía, y así lo hemos vivido juntas. Me lo habéis dado todo. Amistad, cariño, apoyo, risas y llantos, ánimos, visitas, sorpresas... En fin, que las gracias para vosotras no se acabarían nunca.

A Marta, nuevamente, a Assela, Teresa, Bea y Esti, por su apoyo todos estos años y por vuestra gran amistad. Por tantas cenas, tantas vivencias y tantos consejos compartidos.

A mi gente del kayakpolo, especialmente a Carmen, Sonia, Eva, Paloma, Gina y Elena. Sabéis cuánto habéis contribuido en mi felicidad de estos años, en darme fuerza y en propiciar el olvido de toda cuestión que me agobiase relativa a la investigación. Es lo bueno del deporte y de encontrarse con personas como vosotras.

Y a las de balonmano. A las “ahora”, Val, Bárbara, Begoña, Kika, Ana, Pili, Anita, Eva, Noelia, Arantxa, Quirós y Ramón. Además de con vuestra amistad, me habéis ayudado a relajar tensiones, a base de pelotazos... Por otra parte, a las de “antaño” os debo otros tantos millones de gracias. A Ariadna, Susana, Dulce, Mari Carmen, Ruth, Cristina, Regina y Charo, vuestro apoyo continuo, vuestra amistad, todo lo compartido en todos estos años, alegrías y penas, me han ayudado mucho a seguir adelante. También a Inti, cuya amistad y reflexiones valoro mucho.

A mis “guinderos” del BAH, Laura, Jorge, Cristina, Shariffa, César, Paco, Natalia, Jacobo, Bibi, Fer, Pablo, Carlos, Vicen, Edu, Isa y Belinda, que en este último año y medio me habéis enseñado un montón de cosas, a luchar con la palabra y acción y a que las cosas se pueden hacer de otra manera. Hemos compartido mucho más que verduras, y vuestra amistad y energía vital ha sido una gran fuente de ánimos en esta etapa.

A Fulvia, a Jaime, a Nacho, a Jara, a Sara, a David, a Jaane, a Antonio, a Christian, a Henar, a Manolito, a Lesli, a todos los que me habéis apoyado con vuestra amistad y compañía y me habéis dado fuerzas para la tesis y para la vida.

Y finalmente a Noah. POR TODO. Por haber estado siempre ahí, confiando, aguantando, apoyándome y queriéndome. Por su paciencia infinita. Esta tesis es nuestra y nada me hace más feliz que estar escribiendo estas últimas palabras en ellas, por lo que ello representa. Es el mejor regalo que puedo hacerte con mis manos. GRACIAS.

ÍNDICE

RESUMEN	1
SUMMARY	3
1. INTRODUCCIÓN.....	5
1.1. LA HETEROCROMATINA.....	7
1.1.1. La heterocromatina de <i>Drosophila melanogaster</i>	9
1.2. LOS TELÓMEROS.....	15
1.2.1. Los telómeros en <i>D. melanogaster</i>	16
1.3. EL CENTRÓMERO	18
1.3.1. El centrómero en <i>D. melanogaster</i>	20
1.3.1.1. El centrómero del cromosoma Y	21
1.3.1.2. El centrómero del cromosoma 3	22
1.4. SECUENCIACIÓN DE DNA ALTAMENTE REPETIDO	22
2. OBJETIVOS	25
3. MATERIALES Y MÉTODOS	29
3.1. MATERIALES	31
3.1.1. Mantenimiento de <i>D. melanogaster</i> y otras especies de <i>Drosophila</i>	31
3.1.2. Mantenimiento de líneas celulares de <i>D. melanogaster</i>	31
3.1.3. Medios de cultivo bacteriano y soluciones.....	31
3.2. MÉTODOS	32
3.2.1. Preparación del DNA.....	32
3.2.1.1. DNA genómico.....	32
3.2.1.2. DNA de BACs	32
3.2.1.3. DNA de plásmidos	32
3.2.1.4. Extracción del DNA de geles de agarosa.....	32
3.2.2. Marcaje de sondas radioactivas	33
3.2.2.1. Marcaje con $\alpha^{32}\text{P}$ dCTP	33
3.2.2.2. Marcaje con $\gamma^{32}\text{P}$ ATP	33
3.2.3. Rastreo de tres genotecas genómicas de <i>D. melanogaster</i>	33
3.2.4. Digestión del DNA con enzimas de restricción y endonucleasas “homing”	34
3.2.5. Electroforesis en geles de agarosa	34
3.2.5.1. Electroforesis convencional	34
3.2.5.2. Electroforesis de campo pulsado	34
3.2.6. Amplificación de fragmentos de DNA mediante PCR	35
3.2.7. Clonación de productos de PCR o fragmentos de DNA obtenidos mediante digestión enzimática	35

3.2.8. Comprobación de colonias transformantes	36
3.2.8.1. Comprobación de colonias transformantes mediante PCR	36
3.2.8.2. Comprobación de colonias transformantes mediante digestión con enzimas de restricción	36
3.2.9. Transposición <i>in vitro</i>	37
3.2.10. Transformación del DNA	37
3.2.10.1. Transformación mediante choque térmico	37
3.2.10.2. Transformación mediante electroporación	37
3.2.11. Secuenciación del DNA	37
3.2.12. Análisis de la secuencia del DNA	38
3.2.12.1. Búsqueda de secuencias homólogas en bases de datos y comparación de secuencias.	38
3.2.12.2. Ensamblaje de secuencias	38
3.2.13. Hibridación <i>in situ</i> con sondas fluorescentes (FISH)	38
3.2.13.1. Preparación de cromosomas mitóticos	38
3.2.13.2. Marcaje de las sondas	39
3.2.13.3. Condiciones de hibridación	39
3.2.13.4. Inmunodetección de las sondas	40
3.2.14. Inmunofluorescencia e hibridación <i>in situ</i> en fibras de cromatina	40
3.2.14.1. Preparación de fibras de cromatina	40
3.2.14.2. Inmunofluorescencia	41
3.2.14.3. Hibridación <i>in situ</i>	41
 4. RESULTADOS	 43
4.1. EL CENTRÓMERO DEL CROMOSOMA Y DE <i>Drosophila melanogaster</i>	45
4.1.1. Análisis de la región centromérica h18	45
4.1.1.1. Desarrollo de una estrategia de secuenciación para DNA heterocromático	45
4.1.1.2. Obtención del transposón artificial Tn599	45
4.1.1.3. Validación de la nueva técnica de secuenciación	48
4.1.1.4. Obtención de una genoteca del BACR26J21 mediante transposición del transposón Tn599	48
4.1.1.5. Mapeo de la inserción del transposón Tn599	50
4.1.1.6. Comprobación de la aleatoriedad de inserción del transposón en regiones de DNA altamente repetido	54
4.1.1.7. Secuenciación de los clones	55
4.1.1.8. Ensamblado de la secuencia del BACR26J21	55
4.1.1.9. Comparación de las secuencias obtenidas para el BACR26J21 en el CBMSO y el WTSI	57
4.1.1.10. Organización de la región centromérica de h18	58
4.1.1.11. La región centromérica h18 deriva de un telómero	60
4.1.2. Análisis de la región centromérica h17	61
4.1.2.1. Identificación de BACs de la región centromérica h17 del cromosoma Y de <i>D. melanogaster</i>	61
4.1.2.2. La región h17 contiene un gran palíndromo.	65
4.1.2.3. El palíndromo de la región h17 abarca más de 200 kb	67

4.2. EL CENTRÓMERO DEL CROMOSOMA 3 DE <i>D. melanogaster</i>	71
4.2.1. Análisis de la región centromérica del cromosoma 3.....	71
4.2.1.1. La proteína centromérica CID colocaliza con el satélite dodeca en fibras de cromatina.....	71
4.2.1.2. Secuencias relacionadas con el espaciador intergénico ribosomal en h53.....	73
4.2.1.3. El satélite de 10bp está presente en la región proximal de h52.....	78
4.2.1.4. ¿Existe el satélite de 15bp en <i>D. melanogaster</i> ?	81
4.2.1.5. Análisis de la secuencia obtenida de la región centromérica h53	82
5. DISCUSIÓN.....	87
5.1. LA TÉCNICA DE SECUENCIACIÓN	89
5.2. EL CENTRÓMERO EN <i>D. melanogaster</i>	90
6. CONCLUSIONES.....	99
7. BIBLIOGRAFÍA.....	103
8. PUBLICACIONES.....	115
9. ANEXOS	119
ANEXO I. Análisis de los retrotransposones teloméricos presentes en el BACR26J21, a partir de la penúltima repetición de la unidad del satellite 18HT.....	121
ANEXO 2. Análisis de la secuencia de una mitad del palíndromo, presente en el BACR07N15.....	137

ÍNDICE DE FIGURAS

Figura 1. Representación esquemática de la heterocromatina de los cromosomas de <i>D. melanogaster</i> .	10
Figura 2. Localización de los satélites de DNA en la heterocromatina de <i>D. melanogaster</i> .	12
Figura 3. Tipos de elementos transponibles.	13
Figura 4. Esquema de la heterocromatina secuenciada de <i>D. melanogaster</i> .	14
Figura 5. Secuencias teloméricas.	16
Figura 6. Localización del centrómero del cromosoma 3 de <i>D. melanogaster</i> mediante reordenamientos cromosómicos.	22
Figura 7. Plásmido pMOD2<oriV/KAN-2/I-SceI/PI-SceI> y transposón Tn599.	47
Figura 8. Esquema de la técnica de secuenciación mediante trasposición, incorporando la información posicional.	49
Figura 9. Mapeo de la inserción del transposón Tn599 en 15 clones.	51
Figura 10. Mapeo de la inserción del transposón en dos clones en los que el transposón se ha insertado equidistantemente respecto a la diana <i>PI-SceI</i> del vector.	52
Figura 11. Aleatoriedad en la inserción del transposón Tn599 a lo largo del BACR26J21.	54
Figura 12. Distribución de lecturas en el ensamblado de la secuencia del BACR26J21.	57
Figura 13. Deconstrucción de la secuencia del BACR26J21.	60
Figura 14. "Contig" de la región h18 del cromosoma Y, construido mediante "walking".	62
Figura 15. Análisis <i>in silico</i> empleado en la identificación de BACs de la región de inserción del elemento P B783.2.	64
Figura 16. Hibridaciones <i>in situ</i> sobre cromosomas metafásicos.	64
Figura 17. Deconstrucción de la secuencia palindrómica presente en el BACR07N15.	66
Figura 18. "Dotplot" de la primera mitad de la secuencia del BACR07N15 comparada contra la segunda mitad.	66
Figura 19. Análisis de restricción de 6 BACs digeridos con la enzima <i>EcoRI</i> .	69
Figura 20. "Contig" de la región h17 del cromosoma Y.	70
Figura 21. Región centrómerica del cromosoma politénico 3 de <i>D. melanogaster</i> .	72
Figura 22. Cromosomas politénicos del doble mutante <i>Su(var)3-9</i> y <i>SuUR</i> .	72
Figura 23. Hibridación <i>in situ</i> e inmunofluorescencia sobre fibras de cromatina.	73
Figura 24. Localización de <i>Circe</i> , <i>HeT-A</i> y del satélite dodeca en la heterocromatina del cromosoma 3 de <i>D. melanogaster</i> .	74
Figura 25. Hibridaciones <i>in situ</i> sobre cromosomas metafásicos.	76

Figura 26. Hibridación <i>in situ</i> de (a) cromosomas prometafásicos.	77
Figura 27. Hibridación <i>in situ</i> sobre cromosomas politénicos del doble mutante <i>Su(var)3-9</i> y <i>SuUR</i>	78
Figura 28. Hibridación <i>in situ</i> de cromosomas metafásicos.....	79
Figura 29. Hibridaciones en filtro con sondas de los satélites dodeca y 10bp.	80
Figura 30. Hibridaciones <i>in situ</i> sobre cromosomas metafásicos de <i>D. simulans</i> y <i>D. mauritiana</i>	82
Figura 31. Clones procedentes de las genotecas de BACs que contienen el satélite dodeca, ordenados en la regiones h52p- h56.	83
Figura 32. Representación de la anotación de la secuencia obtenida a partir de seis BACs de la región centromérica del cromosoma 3.	85
Figura 33. Tipo de cromosoma Y en las especies del subgrupo melanogaster, en función de la posición del centrómero.....	90
Figura 34. Formación de un gran palíndromo mediante un ciclo de BFB.....	92
Figura 35. Modelo sobre el origen del centrómero.	97

ÍNDICE DE TABLAS

Tabla 1. Satélites de DNA en <i>D. melanogaster</i>	11
Tabla 2. Relación entre el número de clones y la cobertura de lectura de la secuencia.....	50
Tabla 3. Cebadores y condiciones de la PCR para obtener la sonda P.	53
Tabla 4. Cebadores y condiciones de la PCR para obtener la sonda T.	54
Tabla 5. Cebadores y condiciones de la reacción de secuenciación.....	55
Tabla 6. Rango de los subproyectos generados en el ensamblaje de la secuencia del BACR26J21.	56
Tabla 7. Cebadores y condiciones de la PCR para obtener la sonda del pseudo- intrón <i>Mst35B</i>	67
Tabla 8. Clones que hibridaron con la sonda del pseudo-intrón <i>Mst35B</i>	69
Tabla 9. “Contigs” que contienen clones positivos con la sonda del pseudo-intrón <i>Mst35B</i>	69
Tabla 10. Clones que hibridan con la sondas de dodeca y <i>Circe</i> (a), de dodeca y <i>HeT-A</i> (b) y solamente con la sonda de dodeca (c)	74
Tabla 11. Cebadores y condiciones de la PCR para obtener la sonda IGSR.	76
Tabla 12. Secuencia de oligonucleótidos marcados fluorescentemente.	78
Tabla 13. Clones que hibridan con la sonda del satélite 10bp.	81

RESUMEN

La heterocromatina centromérica y telomérica de los cromosomas eucarióticos está compuesta por elementos moderadamente repetidos, como los elementos transponibles y secuencias en tándem altamente repetidas. Debido a esta naturaleza repetida, los Proyectos de Secuenciación de Genomas Completos han fallado a la hora de analizar estas regiones. En esta tesis se describe una estrategia basada en el uso de transposones para secuenciar el DNA altamente repetido, clonado en cromosomas artificiales de bacterias. La clave de esta estrategia consiste en determinar la posición de la inserción del transposón, lo que permite ensamblar correctamente el DNA repetido. Esta estrategia ha sido aplicada a un clon procedente de la región centromérica del cromosoma Y de *Drosophila melanogaster*.

El análisis de la secuencia completa de este clon ha verificado que esta región centromérica evolucionó de un telómero, posiblemente tras una inversión pericéntrica de un cromosoma telocéntrico ancestral. Estos resultados confirman que al utilizar la información posicional en la secuenciación mediante transposones se obtienen mejores resultados que con las estrategias de secuenciación actuales. La técnica descrita puede convertirse en una estrategia universal para descifrar las regiones heterocromáticas de los genomas eucarióticos. Además, se ha caracterizado la región centromérica del cromosoma 3 de *D. melanogaster*. Esta región contiene el satélite 10bp, el satélite dodeca, elementos transponibles y una duplicación segmental del espaciador intergénico ribosomal. Finalmente, se ha visto que la proteína centromérica CID colocaliza con el satélite dodeca en fibras de cromatina.

SUMMARY

The centromeric and telomeric heterochromatin of eukaryotic chromosomes is mainly composed of middle-repetitive elements, such as transposable elements and tandemly repeated DNA sequences. Because of this repetitive nature, Whole Genome Shotgun Projects have failed in sequencing these regions. A transposon-based approach for sequencing highly repetitive DNA sequences in bacterial artificial chromosome clones is described in this thesis. The key to this strategy relies on mapping the precise position of the transposon insertion, which enables the correct assembly of the repeated DNA. This strategy has been applied to a clone from the centromeric region of the Y chromosome of *Drosophila melanogaster*.

The analysis of the complete sequence of this clone has verified that this centromeric region evolved from a telomere, possibly after a pericentric inversion of an ancestral telocentric chromosome. These results confirm that the use of transposon-mediated sequencing including positional mapping information improves current finishing strategies. The technique described could be a universal approach to resolving the heterochromatic regions of eukaryotic genomes. In addition, the centromeric region of the 3rd chromosome of *D. melanogaster* has also been characterized. This region contains the 10bp satellite, the dodeca satellite, transposable elements and a segmental duplication from the ribosomal intergenic spacer. Finally, it has been seen that the centromeric protein CID colocalizes with the dodeca satellite in chromatin fibers.

Introducción

1. INTRODUCCIÓN

1.1. La heterocromatina

Se denomina cromatina al conjunto de DNA y proteínas asociadas al mismo que se encuentran en el núcleo formando los cromosomas. Existen dos tipos de cromatina: la eucromatina y la heterocromatina. En 1928 Heitz describió por primera vez la heterocromatina como los segmentos cromosómicos, o incluso cromosomas enteros, que permanecían condensados durante la interfase de la célula en división (Heitz, 1928). Esta distinción respecto de la eucromatina la realizó a través de tinciones citológicas con colorantes específicos de DNA, en las cuales observó que en el núcleo interfásico unas zonas, las heterocromáticas, se teñían más intensamente que otras. Hoy en día se conocen más características que distinguen a la heterocromatina respecto de la eucromatina. La heterocromatina tiene mayor grado de condensación, replica tardíamente en la fase S, tiene menor densidad de genes, es rica en secuencias repetidas y la histona H3 que la conforma está habitualmente metilada en su lisina 9, mientras que en la eucromatina, la lisina 9 de la histona H3 no aparece metilada, pero sí lo hace la lisina 4 (Grewal y Jia, 2007).

En 1966 Brown realizó una distinción entre dos tipos de heterocromatina: la constitutiva y la facultativa (Brown, 1966). La heterocromatina constitutiva está presente por igual en todas las células del organismo, estando localizada fundamentalmente en el centrómero y la región pericentromérica, en los telómeros y las regiones subteloméricas y en las regiones organizadoras del nucleolo. Por su parte, la heterocromatina facultativa varía su estado de condensación en los distintos tipos celulares y en las diferentes etapas del desarrollo (Grewal y Jia, 2007). Un ejemplo de heterocromatina facultativa se encuentra en el corpúsculo de Barr, como consecuencia de la inactivación de un cromosoma X en las hembras de los mamíferos (Lyon, 1961). El trabajo realizado en esta tesis doctoral se centra en la heterocromatina constitutiva.

Si bien el DNA heterocromático fue considerado durante mucho tiempo como DNA “basura” (Ohno, 1972), debido a su bajo contenido en genes y su inactividad transcripcional, en la actualidad esa visión ha sido modificada

(Makalowski, 2003). En la última década, múltiples grupos han centrado el objeto de estudio en la heterocromatina, a distintos niveles, y en la actualidad se conocen varios procesos en los que la heterocromatina cumple una importante función. Además se han descrito gran número de proteínas asociadas a la heterocromatina, tanto a su formación como mantenimiento, y otro tipo de proteínas implicadas en las modificaciones postraduccionales de las histonas que conforman la heterocromatina. Por todo ello, la heterocromatina ha pasado de un segundo plano, a ocupar un papel muy importante en la comprensión de los mecanismos moleculares que actúan en el cromosoma eucariótico.

Algunas de las funciones claves en las que está implicada la heterocromatina incluyen la regulación de la segregación cromosómica (Allshire et al., 1995; Kellum y Alberts, 1995; Peters et al., 2001a; Peters et al., 2001b), el mantenimiento de la cohesión entre cromátidas hermana (Bernard et al., 2001; Karpen et al., 1996; Wines y Henikoff, 1992) y la protección de la integridad del cromosoma (Cryderman et al., 1999; McCord y Broccoli, 2008). Además, la heterocromatina juega un papel importante en la arquitectura del núcleo interfásico (Dernburg et al., 1996a) y en el apareamiento de los cromosomas homólogos en meiosis (Dernburg et al., 1996b; Karpen et al., 1996; McKee y Karpen, 1990). Una de las funciones de la heterocromatina más estudiadas en la actualidad es la regulación de la expresión génica, a través del silenciamiento génico (Grewal y Jia, 2007). Este fenómeno se descubrió gracias a los estudios de variegación por efecto de posición (PEV) realizados en 1930, por Müller, en los que se observó que cuando un gen eucromático pasaba a localizarse, por un reordenamiento cromosómico o una traslocación, en la heterocromatina, el gen era silenciado y producía un fenotipo variegado (Henikoff, 1990; Muller y Altenburg, 1930; Schultz, 1936). Recientemente, estudios diversos han descrito la presencia de múltiples genes activos en la heterocromatina (Carvalho et al., 2001; Gatti y Pimpinelli, 1992; Smith et al., 2007; Vibranovski et al., 2008). Además, si bien durante mucho tiempo se creyó que la heterocromatina era transcripcionalmente inerte, ahora se sabe que da lugar a RNAs pequeños (siRNA) que, mediante el proceso del RNA interferente (RNAi), dirigen las modificaciones epigenéticas de las histonas y

del DNA (Lippman y Martienssen, 2004). Éstas, a su vez, reprimen epigenéticamente la expresión de genes y de transposones (Brennecke et al., 2007; Elgin y Grewal, 2003). Finalmente, la heterocromatina incluye dos elementos cromosómicos claves en el control del comportamiento cromosómico: el centrómero y el telómero.

A pesar de que la heterocromatina ocupa una fracción considerable del genoma, sobre el 20% del genoma humano y el 35% del de *Drosophila melanogaster*, nuestro entendimiento de la secuencia y organización de la heterocromatina es muy limitado (Hoskins et al., 2007). Esto se debe a su naturaleza altamente repetida. La heterocromatina está compuesta por secuencias cortas repetidas en tándem, denominadas satélites de DNA, por elementos moderadamente repetidos, como los transposones y el DNA ribosomal, y por algunas secuencias de copia única (Weiler y Wakimoto, 1995). Los satélites de DNA son secuencias muy sencillas, y suelen encontrarse en grandes bloques que están a su vez interrumpidos por “islas” de secuencias de mayor complejidad, en su mayoría transposones (Hall et al., 2003). Los transposones, por su parte, no son específicos de la heterocromatina, si bien su presencia en la misma es mucho más elevada que en la eucromatina.

1.1.1. La heterocromatina de *Drosophila melanogaster*

La heterocromatina de *D. melanogaster* comprende 118 Mb de las 351 Mb del genoma de una célula diploide de una mosca hembra, y 139 Mb, en machos. El cromosoma Y entero, el 40% de la parte proximal del cromosoma X, el 25% de los cromosomas 2 y 3, y casi la totalidad del cromosoma 4 son regiones heterocromáticas (Gatti y Pimpinelli, 1992). Esto implica que, alrededor del 35% de la cromatina de *D. melanogaster* está compuesta por heterocromatina.

Las glándulas salivales de larvas de *D. melanogaster* contienen cromosomas politénicos, como resultado de la endoreduplicación. Estos cromosomas están compuestos por gran número de cromátidas homólogas que permanecen juntas. Estos cromosomas tienen un bandeo característico, cuyo patrón fue descrito por Bridges (Bridges, 1935). Los cromosomas politénicos son ampliamente utilizados para la localización citogenética de sondas

euromáticas y para identificar reordenamientos cromosómicos y deleciones. El uso de los estos cromosomas queda limitado a las regiones euromáticas, pues la heterocromatina, debido probablemente a su replicación tardía, permanece agregada en lo que se conoce como el cromocentro. Siendo esto así, Gatti y Pimpinelli realizaron un mapa citológico de la heterocromatina de *D. melanogaster* sobre cromosomas prometáfásicos teñidos con Hoeschst 3358 (Gatti y Pimpinelli, 1992), cuya tinción es similar a la de DAPI. En prometafase los cromosomas todavía no están completamente condensados, de modo que permiten la distinción de las regiones heterocromáticas. De esta forma, Gatti y Pimpinelli identificaron hasta 61 regiones heterocromáticas, a las que nombraron con una “h” seguida por un número. La numeración comenzaba en el brazo largo del cromosoma Y, y continuaba hacia el brazo corto (h1-h25), y así sucesivamente en los cromosomas X (h26-h34), 2 (h35-h46), 3 (h47-h58) y 4 (h59-h61), como muestra la figura 1.

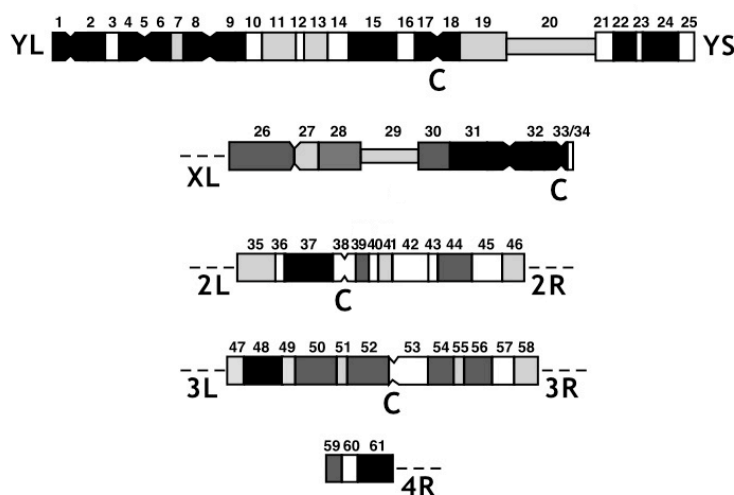


Figura 1. Representación esquemática de la heterocromatina de los cromosomas de *D. melanogaster*.

Los bloques negros, grises y blancos corresponden a regiones de fluorescencia fuerte, media y nula, respectivamente, en la tinción con Hoescht 3358.

La mayor parte de la heterocromatina de *D. melanogaster* está compuesta por satélites de DNA. Estos satélites, que llegan a cubrir megabases de DNA, están formados por repeticiones en tándem de una secuencia básica, que oscila entre 5 y 361 pares de bases (bp, en adelante). Recientemente se han

caracterizado dos nuevos satélites de DNA, el 14HT y el 18HT, cuyas unidades de repetición son de 3,8 kb y de 3,1 kb, respectivamente (Abad et al., 2004a). Entre los bloques de satélite, aparecen regiones de secuencias medianamente repetidas: acúmulos de transposones, truncados en su mayoría y cuya secuencia original ha degenerado. Ocasionalmente, embebidos en la heterocromatina, también se encuentran genes y pseudogenes.

Los satélites de DNA de *D. melanogaster* son redundantes tanto a nivel de secuencia, como en su localización cromosómica. En general, un mismo satélite se encuentra en varias regiones heterocromáticas, si bien alguno, como el satélite dodeca, es específico de la región h53 (ver figura 2, (Lohe et al., 1993). En *D. melanogaster* se han identificado un total de 17 satélites diferentes, que se detallan en la tabla 1.

Tabla 1. Satélites de DNA en *D. melanogaster*

Satélites de DNA en <i>D. melanogaster</i>		
(AATAT) _n	(AATAAAC) _n	rDNA
(AATAC) _n	(AATAGAC) _n	240bp-Rsp
(AATAG) _n	(AAGAGAG) _n	14HT
(AAGAC) _n	(AATAACATAG) _n	18HT
(AAGAG) _n	(ACCGAGTACGGG) _n	Familia 1.688 (260/353/356/359/361bp)
(AACAC) _n	(AACATGTTCTGTTTCG) _n	

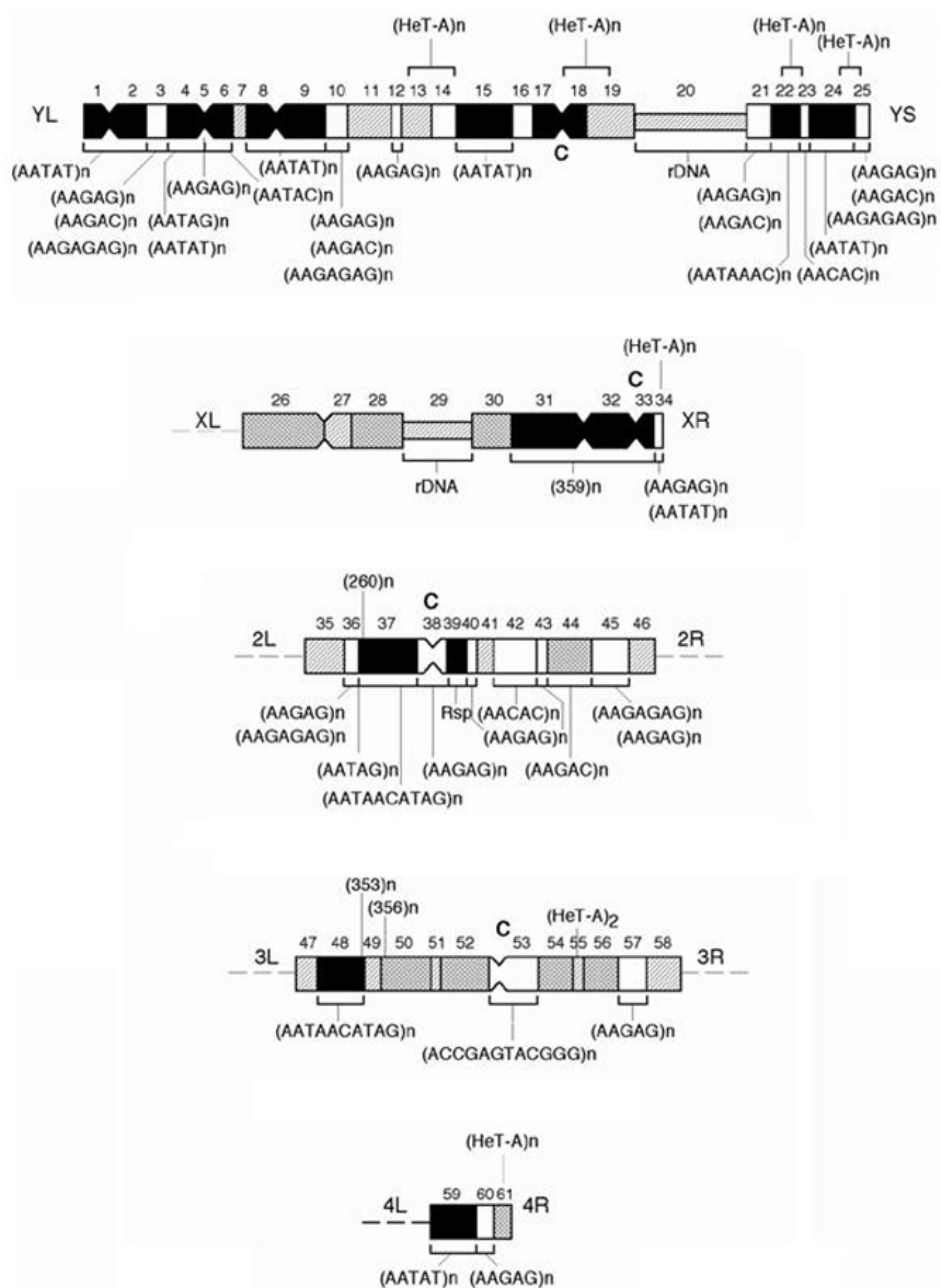


Figura 2. Localización de los satélites de DNA en la heterocromatina de *D. melanogaster*. Modificado de Gatti y Pimpinelli 1992 y Lohe et al. 1993.

Por parte de los elementos transponibles, en *D. melanogaster* se han detectado del orden de 114 familias. Estos se dividen en retrotransposones LTR (“long terminal repeat”, en inglés), retrotransposones no-LTR y transposones TIR (“terminal inverted repeat”, en inglés). Tanto los retrotransposones LTR como los no-LTR se transponen por transcripción inversa, mediante un intermedio de RNA, y por lo tanto, contienen en su secuencia las fases abiertas de lectura (ORF) que codifican para proteínas similares a los genes retrovirales *gag* (proteína con motivos de unión a ácidos nucleicos) y *pol* (transcriptasa inversa). Se distinguen entre sí porque los LTR presentan, en sus extremos, repeticiones terminales; mientras que los no-LTR presentan en su extremo 3’ una secuencia de poliadeninas (poliA). Además, los retrotransposones LTR contienen un tercer ORF (*env*). Por su parte, los transposones TIR se transponen a través de un intermedio de DNA y contienen una ORF que codifica para una transposasa. Éstos se caracterizan por la presencia en sus extremos de repeticiones cortas invertidas. La figura 3 muestra un diagrama de los distintos tipos de elementos transponibles.

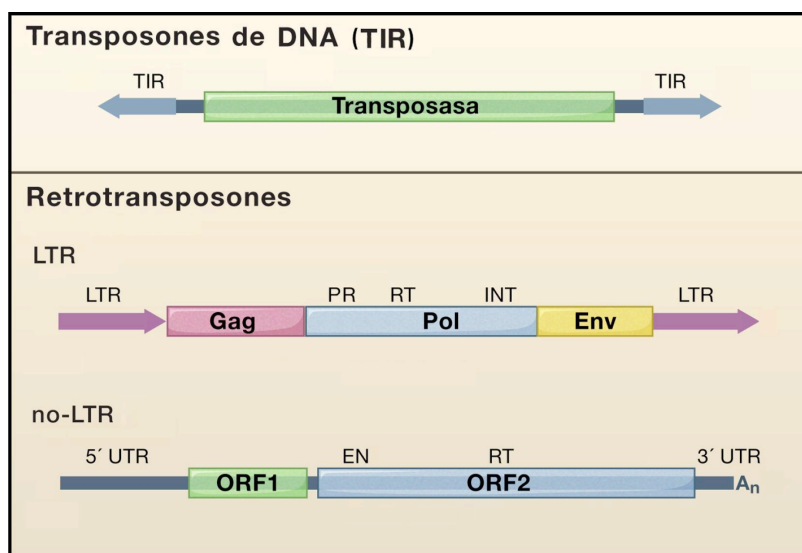


Figura 3. Tipos de elementos transponibles.

Los elementos se dividen en transposones de DNA (TIR) y retrotransposones LTR y no-LTR. Las zonas codificantes de los LTR son *gag*, homóloga a la proteína Gag de retrovirus; *pol*, homóloga a la proteína Pol de retrovirus, y contiene los dominios PR (proteasa), RT (transcriptasa inversa) e INT (integrasa); y *Env*, homóloga a la proteína de la envoltura de retrovirus. Las zonas codificantes de los no-LTR se conocen como ORF1 y ORF2. La ORF2 tiene los dominios EN (endonucleasa) y RT (transcriptasa inversa).

En 2007 Smith y colaboradores publicaron la anotación del “Release 5.1” de la heterocromatina de *D. melanogaster* (Smith et al., 2007). Este trabajo muestra que hasta la fecha tan sólo se han podido secuenciar y anotar correctamente 24 Mb de la porción heterocromática del genoma de *D. melanogaster* (ver figura 4). De esas 24 Mb, alrededor de un 33% correspondían a retrotransposones LTR, otro 33% a elementos no-LTR, un 15% a transposones TIR y tan sólo un 10% a DNA satélite (si bien indican que son conscientes de que esta cifra está infravalorada, debido a la dificultad de clonar el DNA satélite). Esto deja tan sólo un 9% de secuencias únicas, entre las que se encuentran genes que codifican para proteínas (estimados en unos 230 genes) y pseudogenes (32 posibles pseudogenes fueron anotados). La presencia de un número tan elevado de pseudogenes, en comparación con la eucromatina, puede deberse a que, el alto contenido de la heterocromatina en secuencias repetidas sirva como sustrato para recombinaciones, incrementando la frecuencia de duplicaciones en tándem y de duplicaciones segmentales.

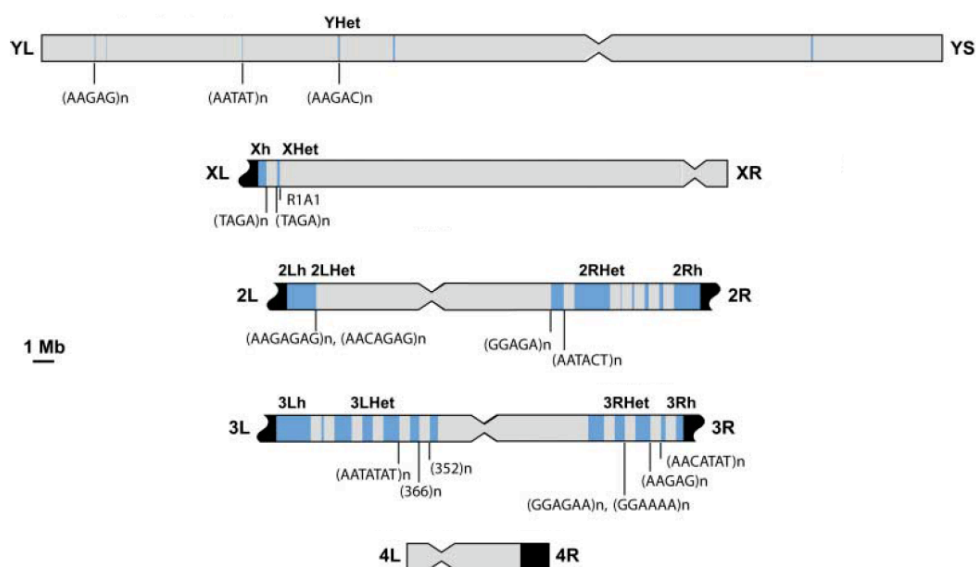


Figura 4. Esquema de la heterocromatina secuenciada de *D. melanogaster*.

En azul se representan las regiones heterocromáticas secuenciadas y ensambladas. El gris representa la heterocromatina no secuenciada y el negro representa el comienzo de la eucromatina.

1.2. Los telómeros

Los telómeros son estructuras nucleoprotéicas localizadas en los extremos de los cromosomas eucarióticos, que fueron detectados al observar distintos comportamientos entre los extremos naturales de los cromosomas y las rupturas del DNA de doble cadena (Muller, 1938). Se han descrito varias funciones en las que los telómeros están implicados. Una de ellas consiste en el mantenimiento de la longitud cromosómica, mediante elongaciones periódicas que contrarrestan la incapacidad de la DNA polimerasa para replicar completamente los cromosomas lineales (Blackburn, 1994; McCord y Broccoli, 2008). Otra de las funciones principales es la de marcar los extremos naturales del cromosoma, de manera que éstos sean distinguidos respecto de los extremos generados por las rupturas de doble cadena, y por lo tanto, no se activen los mecanismos de reparación, degradación o de recombinación (McCord y Broccoli, 2008; Zhu et al., 1999). Además, los telómeros lideran el movimiento de los cromosomas en la profase meiótica (de Lange, 1992; Gilson et al., 1993; Goday y Pimpinelli, 1989; Manzanero y Puertas, 2003; Östergren y Prakken, 1946; Perez et al., 1997; Rhoades y Vilkomerson, 1942), y son importantes para el apareamiento de los cromosomas homólogos (Rockmill y Roeder, 1998). Los telómeros están además involucrados en el envejecimiento de las células, en enfermedades tan importantes como el cáncer y en diversos síndromes con retraso mental (de Lange, 1994; Harley y Villeponteau, 1995; McCord y Broccoli, 2008).

En la mayoría de organismos eucariotas los telómeros están constituidos por repeticiones en tándem de una secuencia corta, de unas 5-8 bp, ricas en guanina y timina, que son añadidas al final de los cromosomas por la telomerasa, una retrotranscriptasa específica de telómeros. En concreto, la secuencia telomérica común a vertebrados, hongos filamentosos y algunos protozoos es (TTAGGG)_n, mientras que en la mayoría de las plantas es (TTTAGGG)_n. Si bien la secuencia de la repetición ha variado a lo largo de la evolución, la riqueza en Gs y su asimetría están conservadas evolutivamente (Meyne et al., 1989). Los telómeros están organizados en estructuras de cromatina de orden superior (Gottschling et al., 1990), que a su vez forman los

lazos-T (“T-loops”) que han sido implicados en la protección de los extremos de los cromosomas (De Lange et al., 1983). Por otra parte, se ha visto que, *in vitro*, la cadena rica en Gs forma estructuras inusuales de DNA con apareamientos no Watson-Crick, denominadas G4-DNA (Sen y Gilbert, 1988; Sundquist y Klug, 1989). El descubrimiento de la proteína POT1, que promueve y estabiliza estas estructuras, favorece la hipótesis de su relevancia *in vivo* (Fang y Cech, 1993).

Es importante distinguir entre las secuencias sintetizadas por la telomerasa (que interaccionan con proteínas específicas de telómero para formar el complejo de protección terminal) y las secuencias subteloméricas TAS (“telomere associated sequences”, en inglés), localizadas a continuación de las repeticiones teloméricas. Las TAS contienen secuencias repetidas, tienen secuencias comunes a varios cromosomas y su longitud puede ser muy variable. También se ha visto que las TAS afectan la expresión de los genes adyacentes (Louis, 1995; Riethman, 2008).

1.2.1. Los telómeros en *D. melanogaster*

En *D. melanogaster*, así como en otras especies de *Drosophila*, la formación de los telómeros no tiene lugar por retrotransposición de cortas repeticiones teloméricas, mediante la telomerasa, sino por retrotransposición de elementos transponibles específicos de telómeros (ver figura 5).

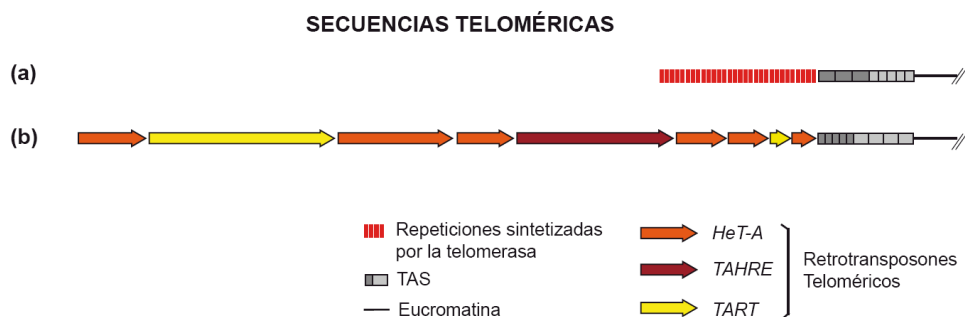


Figura 5. Secuencias teloméricas

(a) sintetizadas por la telomerasa y (b) generadas mediante retrotransposición de transposones teloméricos en *D. melanogaster*.

En el caso de *D. melanogaster* se han descrito tres elementos distintos: *HeT-A* (Danilevskaya et al., 1994; Danilevskaya et al., 1992; Mason y Biessman, 1995; Pardue et al., 1996), *TART* (Levis et al., 1993; Sheen y Levis, 1994) y *TAHRE* (Abad et al., 2004c; Villasante et al., 2007b).

Estos tres elementos son retrotransposones no-LTR que transponen al final de los telómeros, dejando su extremo 3' más próximo al centrómero y el 5' en el final del cromosoma. Esta orientación en la retrotransposición hace que los telómeros de *D. melanogaster* estén formados por tándems de retrotransposones que dan lugar a repeticiones más largas y complejas que las generadas por la telomerasa (Pardue et al., 1997). *HeT-A*, *TART* y *TAHRE* se caracterizan por tener una región 5' UTR -esto es, que no se transcribe ("untranscribed region", en inglés)- generalmente truncada, dado que ésta se encuentra en el extremo expuesto del cromosoma. Estos elementos mantienen la longitud de los telómeros pero no protegen el final del cromosoma del acortamiento causado por el problema de la replicación incompleta (Mason y Biessman, 1995). Además, estos elementos se caracterizan por tener una región 3' UTR especialmente larga, hecho inusual en los retrotransposones. Tanto *TART* como *TAHRE* tienen dos ORFs que codifican para una proteína de unión a ácidos nucleicos y para una retrotranscriptasa. *HeT-A*, sin embargo, carece de la ORF que codifica la retrotranscriptasa.

Recientemente se ha postulado que *HeT-A* deriva del retrotransposón *TAHRE* (con el que comparte gran homología) mediante la pérdida de su ORFII, y que se trataría de un "half-*TAHRE*", hecho que había sido descrito, también, con elementos del tipo LINE (long interspersed nuclear elements) (Consortium, 2004; Smit, 1999) y que se ha observado como un suceso recurrente en varios retrotransposones teloméricos de otras especies de *Drosophila* (Villasante et al., 2007b). Así, el elemento *HeT-A*, carente de retrotranscriptasa, se transpone gracias a una retrotranscriptasa suministrada en *trans*, probablemente a partir del elemento completo *TAHRE*, como sucede con los elementos LINE y half-LINE (Wei et al., 2001).

Por último es importante mencionar que, si bien los retrotransposones que se han descrito se definen como teloméricos, se han hallado secuencias de los mismos presentes en regiones centroméricas y otras regiones heterocromáticas, pero nunca en la eucromatina (Abad et al., 2004a; Agudo et al., 1999; Berloco et al., 2005; Danilevskaya et al., 1993; Losada et al., 1999a; Losada et al., 1997; Traverse y Pardue, 1989).

1.3. El centrómero

La región del cromosoma especializada en dirigir la segregación de las cromátidas hermanas a las células hijas durante la mitosis se conoce como el centrómero. Este aparece citológicamente visible como una constricción primaria en los cromosomas mitóticos de los eucariotas superiores. En el centrómero es donde se ensambla el cinetocoro, una estructura proteica compleja, que interacciona con los microtúbulos durante la segregación cromosómica.

Existen dos clases principales de centrómeros en eucariotas: los centrómeros difusos y los centrómeros localizados (Ekwall, 2007). Los centrómeros localizados se dividen, a su vez, en dos modalidades: puntuales y regionales. Los centrómeros difusos, como los presentes en los cromosomas holocéntricos de *Caenorhabditis elegans*, se unen a los microtúbulos mediante determinantes cinetocóricos distribuidos a lo largo del cromosoma (Maddox et al., 2004). En los centrómeros puntuales la formación del cinetocoro se produce en una zona discreta especificada por una secuencia concreta (Pluta et al., 1995). Así, en *Saccharomyces cerevisiae*, el DNA centromérico está formado por una secuencia de 125 bp que contiene tres subregiones: los elementos CDEI, CDEII y CDEIII (Clarke y Carbon, 1980; Clarke y Carbon, 1985). El cinetocoro de los cromosomas puntuales se une a un solo microtúbulo. En los centrómeros regionales, mucho más comunes (*Schizosaccharomyces pombe*, *Drosophila melanogaster*, *Homo sapiens*), el DNA centromérico contiene distintas repeticiones en tándem y el cinetocoro interacciona con varios microtúbulos. En *S. Pombe* el centrómero está formado por un elemento central (ctr) de 4-7 kb, que está flaqueado por las secuencias “imr” y por las

repeticiones “otr” que llegan a cubrir hasta 110 kb (Clarke, 1990; Clarke et al., 1986).

En eucariotas superiores el DNA centromérico se caracteriza por su contenido en secuencias en tándem altamente repetidas o satélites de DNA. En particular, en humanos se ha visto que todos sus centrómeros tienen bloques del denominado satélite alfoide (de hasta 7 Mb) con unidades de 171 bp orientadas cabeza-a-cola (Choo et al., 1991; Maio, 1971; Mitchel et al., 1985). La secuencia de estas repeticiones es globalmente rica en AT pero contienen un motivo de 17 bp, denominado “CENP-B box”, rico en GC, que funciona como sitio de unión a la proteína centromérica CENP-B (Earnshaw y Rothfield, 1985; Earnshaw et al., 1987; Masumoto et al., 1989). En las regiones más externas de los bloques de satélite alfoide, éste está interrumpido por secuencias SINE (short intersperesed nuclear elements) y LINE (Prades et al., 1996).

A pesar de la alta conservación que existe en la maquinaria responsable de la segregación cromosómica en los eucariotas, la naturaleza repetida de las secuencias centroméricas hace que éstas evolucionen rápidamente e impiden que estén conservadas. Debido a la dificultad que existe para analizar, clonar, y secuenciar secuencias repetidas, actualmente no se conoce la secuencia completa de la mayoría de los centrómeros (Hoskins et al., 2007; Rudd y Willard, 2004; Schueler et al., 2001).

Por otra parte, estudios sobre neocentrómeros, han mostrado que, en ocasiones, los centrómeros pueden formarse en regiones que carecen de las secuencias centroméricas conocidas, y que, sin embargo, forman un cinetocoro que se propaga a las siguientes generaciones. Esto hizo pensar que se requiere una conformación especial de la heterocromatina, la cual, mediante mecanismos epigenéticos, controla la formación y el mantenimiento del centrómero funcional (Vagnarelli et al., 2008).

La cromatina de todos los centrómeros (incluidos los neocentrómeros) contienen una variante de la histona H3: la proteína centromérica CENP-A

(Harrington et al., 1997; Voullaire et al., 1993). La proteína CENP-A se utiliza rutinariamente como marcador centromérico dado que se observó que en células deficientes en CENP-A se bloquea la formación del centrómero (Amor et al., 2004; Goshima et al., 2003; Howman et al., 2000). La cromatina centromérica está formada por zonas con nucleosomas que contienen CENP-A entremezcladas con zonas donde los nucleosomas contienen la histona H3, en lugar de CENP-A (Blower et al., 2002). Además, en estas zonas se ha observado que la histona H3 aparece dimetilada en su lisina 4 (Sullivan y Karpen, 2004). Curiosamente, esta metilación de la histona H3 es característica de las regiones eucromáticas (Jenuwein y Allis, 2001; Kouzarides, 2007), cuando normalmente las regiones heterocromáticas presentan metilación de la lisina 9 de la histona H3.

Si bien las explicaciones sobre los mecanismos que subyacen la plasticidad de los centrómeros (a nivel de secuencia) se han centrado en la epigenesis (Karpen y Allshire, 1997), la existencia de RNAs codificados por los satélites centroméricos (Hall et al., 2002; Volpe et al., 2002) podría también explicar cómo las repeticiones centroméricas evolucionan tan rápidamente, manteniendo su función (Topp et al., 2004). De hecho, Wong y colaboradores han sugerido que los transcritos de los satélites centroméricos podrían facilitar el ensamblaje de nucleoproteínas específicas del centrómero (Wong et al., 2007).

1.3.1. El centrómero en *D. melanogaster*

Estudios citogenéticos permitieron a Lohe y colaboradores localizar los satélites de DNA presentes en el genoma de *D. melanogaster*, como se muestra en la figura 2 (Lohe et al., 1993). Estos estudios muestran que la mayoría de los satélites se encuentran en varias localizaciones no centroméricas. Además, tampoco se detectaron satélites comunes a todos los centrómeros (Carmena y Gonzalez, 1995; Pimpinelli et al., 1994). El estudio realizado sobre el centrómero del minicromosoma Dp1187, derivado del cromosoma X, delimitó la función centromérica a 420 kb (Murphy y Karpen, 1995; Sun et al., 1997). En esa región se describió la presencia de bloques de los satélites de DNA

(AAGAG)_n y (AATAT)_n. Éstos satélites no son específicos de centrómeros (ver figura 2). Sin embargo, estudios sobre otros cromosomas han detectado satélites centroméricos específicos: en el cromosoma Y se ha descrito el satélite 18HT (Abad et al., 2004a; Agudo et al., 1999), mientras que en el cromosoma 3 se ha descrito el satélite dodeca (Abad et al., 1992). Según esto, en *D. melanogaster* el centrómero de cada cromosoma parece tener una organización estructural diferente pero semejante a los del resto de los eucariotas superiores. Esto hace de los cromosomas de *D. melanogaster* un buen sistema modelo para el estudio de la heterocromatina centromérica. Sin embargo, para poder dilucidar cómo se determina la identidad centromérica, y la relevancia de la composición del DNA centromérico, es preciso determinar la estructura de varios centrómeros. Por ello, este trabajo se ha centrado en el análisis estructural del centrómero de dos cromosomas de *D. melanogaster*: el cromosoma 3 y el cromosoma Y.

1.3.1.1. El centrómero del cromosoma Y

La presencia de retrotransposones específicos de telómeros, como *HeT-A* y *TART*, en la región centromérica h18 del centrómero del cromosoma Y de *D. melanogaster* viene siendo objeto de estudio durante la última década (Abad et al., 2004a; Agudo et al., 1999). Hasta la fecha dos hipótesis han sido propuestas: una defiende la idea de que pequeños fragmentos de los telómeros podrían haberse movido a regiones internas del cromosoma Y (Danilevskaya et al., 1993; George et al., 2006), mientras que la otra hipótesis propone que es un telómero entero o casi entero quien está presente en el centrómero del cromosoma Y (Abad et al., 2004a; Agudo et al., 1999). Los dos posibles escenarios se asemejan, siendo el tamaño del fragmento presente en el centrómero la única diferencia aparente. Sin embargo, los mecanismos que habrían originado cada una de esas situaciones son muy distintos. Como bien se conoce, durante la evolución los cromosomas Y adquieren duplicaciones segmentales con cierta frecuencia (Gvozdev et al., 2005). En consecuencia, se podría pensar que los fragmentos teloméricos encontrados en regiones no teloméricas del cromosoma Y de *Drosophila* podrían ser duplicaciones segmentales de distintos telómeros. Por otra parte, si se encontrasen evidencias de la presencia de un telómero entero en el centrómero del

cromosoma Y, entonces cabría pensar que el origen más probable sea un reordenamiento intra-cromosómico. Con el objetivo de resolver esta cuestión sería necesario determinar la secuencia de gran parte de la región h18.

1.3.1.2. El centrómero del cromosoma 3

Koryakov y colaboradores localizaron, mediante reordenamientos cromosómicos, el centrómero del cromosoma 3 de *D. melanogaster* en la banda heterocromática 53 (h53), coincidiendo con la localización de la constricción primaria observada en cromosomas prometafásicos (ver figura 6) (Koryakov et al., 2002).

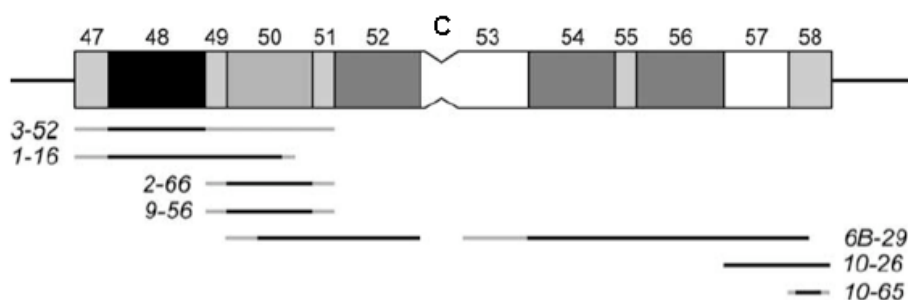


Figura 6. Localización del centrómero del cromosoma 3 de *D. melanogaster* mediante reordenamientos cromosómicos.

Las líneas representan las deficiencias estudiadas. Extraído de Koryakov et al. (2002)

En esta región se ha construido un mapa físico de, aproximadamente, 1,5 Mb que muestra dos bloques principales del satélite dodeca (Losada et al., 2000). Este satélite está formado por la repetición en tándem de unidades de 11 bp (ACCAGTACGGG) ó de 12 bp (ACCGAGTACGGG) (Abad et al., 1992) y ha sido localizado, también, en los centrómeros de los cromosomas 2 y 3 de *D. simulans* y *D. mauritiana* (Carmena et al., 1993). Además, el satélite dodeca parece estar conservado evolutivamente, pues muestra hibridación cruzada con secuencias presentes en los centrómeros de distintas especies (Abad et al., 1992).

1.4. Secuenciación de DNA altamente repetido

La heterocromatina centromérica y telomérica, como ya se ha señalado, está principalmente compuesta por elementos moderadamente repetidos y por satélites de DNA. Dado que estas regiones tienen funciones importantes en la

segregación cromosómica y en la arquitectura nuclear, es necesaria una descripción detallada de su secuencia para lograr entender el comportamiento cromosómico y la evolución de los cromosomas.

El método de secuenciación más utilizado en los proyectos de secuenciación de genomas completos, es el que se conoce como la técnica “shotgun”. En esta técnica el DNA genómico de la especie en cuestión es fragmentado y clonado para generar una genoteca con insertos de ~2 kb. Los clones de la genoteca son secuenciados y las lecturas son ensambladas en base a solapamientos. Habitualmente se genera también una genoteca de cromosomas artificiales de bacterias (BACs, en adelante) y a partir de esta genoteca se obtiene el número mínimo de BACs que cubre el genoma completo. Normalmente, sólo se secuencian los clones de esta genoteca que corresponden a regiones que no han podido ser descifradas mediante el “shotgun” inicial.

La naturaleza repetida de la heterocromatina dificulta su clonación, mapeo, secuenciación y ensamblado. Además, el DNA repetido muestra un alto grado de inestabilidad en clones de bajo número de copias (plásmidos). Por este motivo para clonar grandes fragmentos de DNA genómico se utilizan BACs, que al replicarse en bajo número de copias, proporcionan mayor estabilidad.

El principal problema de la secuenciación de la heterocromatina tiene lugar durante el proceso de ensamblado, cuando aparecen lecturas de secuencias idénticas o casi idénticas, y el programa de ensamblado, basado en solapamientos de secuencia, no es capaz de distinguirlas.

Como consecuencia de estas dificultades técnicas, la heterocromatina ha sido excluida de la mayoría de los grandes proyectos de secuenciación (Myers et al., 2000; Venter et al., 2001), en los que el objetivo primordial era obtener la secuencia de la porción eucromática del genoma. Por lo tanto, si se quiere conocer la estructura de la heterocromatina de los organismos eucariotas superiores, es necesario desarrollar nuevas técnicas de secuenciación, que permitan el análisis del DNA altamente repetido.

Objetivos

2. OBJETIVOS

El estudio realizado en esta tesis doctoral tiene como meta final encontrar los determinantes estructurales de la función centromérica en el cromosoma eucariótico. Para ello, se han utilizado los cromosomas 3 e Y de *Drosophila melanogaster* como modelo de estudio. Los objetivos concretos que se persiguen en el presente trabajo son los siguientes:

1. Desarrollar una nueva técnica de secuenciación que permita obtener fácilmente la secuencia de las regiones heterocromáticas.
2. Determinar el origen de la región centromérica h18 del cromosoma Y de *D. melanogaster*.
3. Caracterizar la región h17 del cromosoma Y de *D. melanogaster*.
4. Determinar la estructura de la región centromérica h53 del cromosoma 3 de *D. melanogaster*.

Materiales y Métodos

3. MATERIALES Y MÉTODOS

3.1. Materiales

3.1.1. Mantenimiento de *D. melanogaster* y otras especies de *Drosophila*

El organismo modelo empleado en esta tesis doctoral para abordar el estudio de las regiones centroméricas de los cromosomas eucariontes es *Drosophila melanogaster*. Como cepa silvestre de *D. melanogaster* se ha usado Vallecas. El mapeo físico de los satélites centroméricos del cromosoma 3 se ha realizado en la cepa isogénica *red* e (ésta cepa isogénica para el cromosoma 3 fue construida por el Prof. Ripoll en 1986). En el estudio de conservación de satélites centroméricos también fueron empleadas otras especies, del subgrupo melanogaster, como *D. mauritiana* y *D. simulans*. Todas estas especies fueron adquiridas a través del “Bloomington Drosophila Stock Center”, de la Universidad de Indiana, Estados Unidos. El mantenimiento de las distintas especies se realizó a una temperatura de 18°C, en botellas de vidrio que contenían una papilla a base de harina, levadura, glucosa, agar, nipagin y ácido propiónico como fuente de alimento.

3.1.2. Mantenimiento de líneas celulares de *D. melanogaster*

En el presente trabajo se emplearon células S2 de *D. melanogaster*, las cuales se crecieron a 20°C en un medio de cultivo compuesto por medio Schneider complementado con suero bovino fetal inactivado por calor (10%), penicilina y estreptomicina (100 µg/ml).

3.1.3. Medios de cultivo bacteriano y soluciones

La composición de los medios de cultivo y soluciones utilizadas en esta Tesis se encuentran descritas en (Sambrook et al., 1989).

3.2. Métodos

3.2.1. Preparación del DNA

3.2.1.1. DNA genómico

Para la extracción del DNA genómico de *D. melanogaster*, *D. simulans* y *D. mauritiana* se utilizaron moscas adultas tal y como se describe en (Pirrotta et al., 1983). La preparación del DNA genómico de alto peso molecular a partir de embriones de 0 a 12 horas se realizó en bloques de agarosa según se describe en (Abad et al., 1992).

3.2.1.2. DNA de BACs

En función de las necesidades de cantidad de DNA a obtener, la extracción del DNA de los BACs se realizó de distinta manera: en las maxi-preparaciones se partió de 500 ml de cultivo bacteriano (crecido 16 horas con agitación sobre medio LB con 12,5 µg/ml de cloranfenicol) y se procedió a extraer el DNA mediante lisis alcalina, seguida de una purificación mediante precipitación con polietilenglicol, según queda descrito en (Sambrook et al., 1989). En el caso de mini-preparaciones, se partió de 10 ml de cultivo bacteriano, y tras la lisis alcalina se purificó el DNA mediante extracción con fenol: cloroformo: isoamilico (25:24:1) y posteriormente se precipitó el DNA con isopropanol. Finalmente, en la extracción del DNA de los clones de la genoteca obtenida por transposición, que iban a ser secuenciados, se empleó el kit para extracción de DNA de BACs en placas de 96 pocillos de Quiagen, siguiendo las especificaciones de la casa comercial. (R.E.A.L Prep 96 Plasmid Kit, Quiagen).

3.2.1.3. DNA de plásmidos

El DNA de plásmido se obtuvo empleando el kit “Wizard Plus Minipreps DNA Purification Systems” (Promega), siguiendo las especificaciones de la casa comercial.

3.2.1.4. Extracción del DNA de geles de agarosa

Para extraer el DNA de geles de agarosa, se cortó la banda de interés y se empleó el kit “Wizard® SV Gel and PCR Clean-Up System” (Promega).

3.2.2. Marcaje de sondas radioactivas

3.2.2.1. Marcaje con $\alpha^{32}\text{P}$ dCTP

Las sondas P, T, Mst35Ba y dodeca se marcaron radioactivamente con $\alpha^{32}\text{P}$ dCTP, partiendo del DNA amplificado mediante PCR, siguiendo las instrucciones del kit “Rediprime Random Prime Labelling System” (Amersham Pharmacia Biotech).

3.2.2.2. Marcaje con $\gamma^{32}\text{P}$ ATP

La sonda 10bp se marcó radioactivamente con $\gamma^{32}\text{P}$ ATP. Se marcaron 100 ng del oligonucleótido 10bp, con 10 μCi de $\gamma^{32}\text{P}$ ATP (Perkin Elmer), 20 U de T4 PNK en un medio que contenía 500 mM Tris-HCl (pH 7.6 a 25°C), 100 mM MgCl_2 , 50 mM DTT, 1 mM espermidina y 1 mM EDTA. La incubación se llevó a cabo durante una hora a 37°C.

3.2.3. Rastreo de tres genotecas genómicas de *D. melanogaster*

Los filtros de las genotecas de BACs de *D. melanogaster* RPCI-98, CHORI221 y CHORI223 fueron rastreados con las sondas Mst35Ba, dodeca y 10bp, en busca de clones de regiones centroméricas de los cromosomas 3 e Y de *D. melanogaster*.

En el caso de las sondas Mst35Ba y dodeca, los filtros se prehibridaron a 65°C durante una hora en una solución que contiene 10% BSA, 0,5M EDTA, 1M $\text{NaH}_2\text{PO}_4\text{-Na}_2\text{HPO}_4$ y 20% SDS. Los filtros se hibridaron 16 horas a 65°C en la misma solución con las sondas marcadas radioactivamente y desnaturalizada 5 minutos a 95°C. Se realizaron dos lavados de 30 minutos a 65°C en 2x SSC y 0,1% SDS seguidos por dos lavados de 15 minutos a 65°C en 0,2x SSC y 0,1% SDS.

En el caso de la sonda 10bp, por su alto contenido en A y T, se redujo la temperatura de prehibridación e hibridación y los lavados posteriores se realizaron a una temperatura de 50°C.

Los filtros fueron expuestos a películas de autorradiografía Kodak, a -80°C , con pantallas amplificadoras de señal, durante el tiempo necesario para detectar la señal, que normalmente fue de una noche.

3.2.4. Digestión del DNA con enzimas de restricción y endonucleasas “homing”

En múltiples ocasiones a lo largo del desarrollo de este trabajo se digirió DNA de distinta procedencia con enzimas de restricción siguiendo, en cada caso, las instrucciones especificadas por el fabricante (NEB). En el caso de las endonucleasas “homing”, las digestiones se llevaron a cabo durante 4 horas a 37°C , fueron inactivadas mediante calentamiento durante 10 minutos a 65°C y se adicionó 1 μl de 10% SDS. Los fragmentos de restricción se corrieron en geles de agarosa, como se describe en el siguiente apartado.

3.2.5. Electroforesis en geles de agarosa

3.2.5.1. Electroforesis convencional

Los DNAs genómicos, de BACs, de plásmidos y los DNAs obtenidos mediante PCR se corrieron en geles de agarosa de concentraciones comprendidas entre el 0.8% y 1% en tampón 0,5x TAE (20 mM Tris-acetate y 0,5 mM EDTA, pH 8.0). Como marcadores de peso molecular se usaron múltiplos del DNA del fago λ , el DNA del fago λ digerido con HindIII y el DNA del fago $\phi 29$ digerido con HindIII. La electroforesis se llevó a cabo siguiendo el protocolo descrito en (Sambrook et al., 1989). Los geles de agarosa fueron teñidos con bromuro de etidio (0,5 mg/ml), lavados en agua destilada y fotografiados bajo un transiluminador de luz blanca/UV acoplado a una cámara (UVP Laboratory Products).

3.2.5.2. Electroforesis de campo pulsado

Cuando la digestión del DNA de BACs producía fragmentos de gran tamaño, éstos se separaban mediante electroforesis de campo pulsado en geles de agarosa del 1% en tampón 0,5x TAE (20 mM acetato de Tris y 0,5 mM EDTA, pH 8.0). La electroforesis se llevó a cabo en un aparato CHEF-DRII (Bio-Rad) a

6 V/cm, con un gradiente de pulso de 0,4 a 10 segundos, durante 20 horas a 14°C. Como marcadores de peso molecular se usaron los marcadores comerciales Low Range PFG y Mid Range PFG (NEB). Los geles de agarosa fueron teñidos con bromuro de etidio (0,5 mg/ml), lavados en agua destilada y fotografiados bajo un transiluminador de luz ultravioleta. Los fragmentos de alto peso molecular obtenidos tras la digestión del DNA de embriones se analizaron utilizando el aparato de electroforesis de campo pulsado Waltzer, desarrollado por (Southern et al., 1987), pues la resolución de éste es mayor que la del CHEF-DRII.

3.2.6. Amplificación de fragmentos de DNA mediante PCR

Las concentraciones de los componentes de las reacciones de PCR utilizadas en las amplificaciones realizadas en esta tesis doctoral, fueron estándar: 0.2 mM dNTPs, 1.5 mM MgCl₂, 0.1 µM de cada uno de los cebadores y 2 unidades de DNA Taq polimerasa (Promega) en un volumen final de 50 µl. En ocasiones se emplearon distintas polimerasas, como la DNA polimerasa “FideliTaqTM PCR Master Mix” (usb), para amplificar el pseudogen Mst35Ba, y las polimerasas Expand Long Template PCR System (Roche) y LA PCR Kit (donde el “LA” representa “Long and accurate”) para amplificar fragmentos de tamaño superior a 5 kb. Las reacciones de PCR se llevaron a cabo en un termociclador Perkin-Elmer 9600 y en un Perkin Elmer 2400. Los programas de PCR y las parejas de cebadores usados en cada reacción se especifican en el apartado correspondiente.

3.2.7. Clonación de productos de PCR o fragmentos de DNA obtenidos mediante digestión enzimática

Las distintas clonaciones realizadas en el transcurso de esta tesis doctoral, se realizaron sobre distintos vectores, algunos comerciales, como pBluescript II KS- (Stratagene) y pMOD2<oriV/KAN-2> (Epicentre), y otros fueron construcciones previamente obtenidas. Los insertos se obtuvieron o bien por PCR o bien mediante digestión enzimática. En los casos necesarios tanto vector como inserto fueron incubados con una mezcla de dideoxinúcleótidos (0,1 mM) y las enzimas T4 DNA polimerasa y Klenow (Roche), en una solución que contenía 50 mM Tris-HCl, 15 mM (NH₄)SO₄, 7 mM MgCl₂, 0.1 mM EDTA,

10 mM 2-mercaptoethanol y 0.02 mg/ml BSA, para generar extremos romos. Las incubaciones se hicieron primero sólo con una unidad de enzima T4 DNA polimerasa, 40 minutos a temperatura ambiente y 10 minutos a 37°C, y posteriormente se adicionaron dos unidades de enzima klenow y se incubó otros 30 minutos a 37°C. Los vectores fueron siempre defosforilados, con dos unidades de fostatasa alcalina (Roche) en 50 mM Tris-HCl y 0,1 mM EDTA durante cuatro periodos de 15 minutos alternando las temperaturas de 37°C y 56°C. Las ligaciones se realizaron mediante el “Fast-Link™ DNA Ligation Kit” (Epicentre) siguiendo las indicaciones del proveedor. Se transformó mediante choque térmico en células competentes DH5 α o DH10B, según las necesidades de la construcción en cuestión. Los detalles particulares de cada clonación se detallan en el apartado correspondiente.

3.2.8. Comprobación de colonias transformantes

3.2.8.1. Comprobación de colonias transformantes mediante PCR

Cuando se contaba con una pareja de cebadores a ambos extremos del inserto, o incluso en zonas interiores, se comprobaba la presencia del mismo mediante PCR. En lugar de extraer el DNA de cada colonia, se pinchaba directamente cada colonia con un palillo, que se introducía en la mezcla de reacción. Durante la desnaturalización inicial de la PCR es de suponer que las células se rompen por el calor, dejando el DNA más accesible. Tras la PCR, 5 μ l de cada amplificación se cargaban en un gel de agarosa, y los fragmentos de DNA se separaban mediante una electroforesis convencional. Posteriormente las moléculas de DNA eran visualizadas mediante luz ultravioleta.

3.2.8.2. Comprobación de colonias transformantes mediante digestión con enzimas de restricción

En ocasiones, la presencia de inserto se analizó mediante digestiones enzimáticas. Así, observando el incremento del tamaño de una o varias bandas, o la aparición de bandas nuevas es posible detectar la presencia del inserto e incluso la orientación del mismo. Las enzimas de restricción utilizadas en cada caso, serán detalladas en el apartado correspondiente.

3.2.9. Transposición *in-vitro*

Se formaron “transposomas” incubando 1 µg del transposón junto con 10 µg/µl de transposasa hiperactiva, en una solución que contenía 50 mM de tris-acetato (pH 7.5), 150 mM de acetato potásico, 10 mM de acetato magnésico y 4 mM de espermidina. Las incubaciones para la formación de los complejos se llevaron a cabo durante 2 horas a 37°C en un volumen total de 40 µl. Para la transposición *in-vitro* se incubaron 14 µl de los “transposomas” con 1 µg del DNA de interés, durante 2 horas a 37°C, en un volumen de reacción de 20 µl. Las reacciones se pararon con 0.1% SDS y 10 minutos de incubación a 70°C.

3.2.10. Transformación del DNA

3.2.10.1. Transformación mediante choque térmico

La transformación mediante choque térmico se empleó para introducir en las células plásmidos de tamaño inferior a 10 kb y se realizó tal como está descrito en (Sambrook et al., 1989). Para las transformaciones se utilizaron células competentes DH5α o DH10B, obtenidas en el servicio de fermentación del CBMSO.

3.2.10.2. Transformación mediante electroporación

La transformación mediante electroporación se empleó para introducir en las células plásmidos de tamaño superior a 10 kb y DNA de BACs. Las electroporaciones se realizaron en un equipo de electroporación Easyjet Plus EquiBio, con cubetas de 0,2 cm de grosor, y bajo las siguientes condiciones: 5 ms, 2.5 Kv, 25 µF, 200 ohmnios. Las células empleadas fueron TransforMax™ EPI300™ Electrocompetent *E. coli* (Epicentre).

3.2.11. Secuenciación del DNA

La secuenciación del DNA se realizó en la Unidad de Genómica “Antonia Martín Gallardo” del Parque Científico de Madrid, PCM-UAM, mediante la química de terminadores marcados (Big Dye Terminators, v3.1, Applied Biosystems). Las condiciones de la reacción de secuenciación empleadas fueron las recomendadas por la casa comercial (Applied Biosystems). Las reacciones de secuenciación se analizaron en un secuenciador ABI Prism

3730. Los cebadores empleados para cada reacción de secuenciación se detallan en el apartado correspondiente.

3.2.12. Análisis de la secuencia del DNA

3.2.12.1. Búsqueda de secuencias homólogas en bases de datos y comparación de secuencias.

Las secuencias obtenidas fueron analizadas empleando el programa informático BLAST, que las compara contra todas las secuencias existentes en una base de datos determinada. Las bases de datos empleadas fueron las de Flybase (www.flybase.org) y del Genbank (<http://BLAST.ncbi.nlm.nih.gov/>).

Las herramientas utilizadas para comparar dos secuencias entre sí fueron BLAST2 (<http://BLAST.ncbi.nlm.nih.gov/bl2seq/wBLAST2.cgi>), y “Dotplot” (<http://www.vivo.colostate.edu/molkit/dnadot/>).

3.2.12.2. Ensamblaje de secuencias

Las lecturas fueron ensambladas utilizando los programas del paquete Staden, que incluyen los programas preGap4, Gap4, Trev y Spin (Bonfield et al., 1995; Staden, 1996). Los detalles del ensamblaje de las secuencias del BACR26J21 se detallan en el apartado correspondiente.

3.2.13. Hibridación *in situ* con sondas fluorescentes (FISH)

3.2.13.1. Preparación de cromosomas mitóticos

Los cromosomas prometafásicos fueron obtenidos a partir de neuroBLASTos de larvas en el tercer estadio. La extracción de los cerebros se llevó a cabo en solución salina (1x PBS). Los cerebros fueron sumergidos durante 5 minutos en citrato sódico 0,5 M, transferidos a una gota de ácido acético al 45% sobre un cubreobjeto recubierto con silicona, donde permanecieron durante 90 segundos, pasados los cuales se eliminó el ácido acético al 45% y se añadieron 2 µl de ácido acético al 60% a cada cerebro. Entonces se procedió a realizar el aplastado y a congelar las preparaciones en nitrógeno líquido. Tras la congelación las preparaciones se deshidrataron en etanol absoluto a -20°C, donde se conservaron hasta su posterior utilización.

3.2.13.2. Marcaje de las sondas

En el transcurso de esta tesis doctoral se han empleado distintos tipos de sondas y de marcaje de las mismas. Por un lado, para hibridar DNA satélite, se emplearon oligonucleótidos marcados en origen, por la propia casa comercial (New England Biolabs). Éste fue el caso de las sondas:

- Dodeca, oligonucleótido de 46 nt, que fue marcado con fluoresceína en el extremo 3' y con cy3 en el extremo 5',
- 10bp, oligonucleótido de 50 nt, que fue marcado con fluoresceína en el extremo 3' y con cy3 en el extremo 5',
- 15bp, oligonucleótido de 41 nt, que fue marcado con cy3 en el extremo 5',

Por otro lado, tanto el DNA de los BACs de la genoteca RPCI-98 de *D. melanogaster* como los DNAs amplificados mediante PCR fueron marcados con digoxigenina mediante “nick translation” siguiendo las indicaciones del kit “DIG-Nick Translation Mix” de la casa comercial Roche. Estas sondas, tras ser marcadas fueron precipitadas con acetato sódico (3M) y etanol absoluto, y guardadas a -20°C hasta su uso posterior.

3.2.13.3. Condiciones de hibridación

El DNA de las preparaciones se desnaturalizó en formamida al 70%, 2x SSC a 70°C durante 2 minutos, tras lo cual las preparaciones se deshidrataron en etanol al 75%, al 90% y absoluto, todas ellas a -20°C, durante 5 minutos cada paso y se secaron al aire.

Las sondas fueron disueltas en la solución de hibridación (50% formamida, 10% sulfato de dextrano y 2x SSC). Para cada preparación se emplearon alrededor de 250 ng de DNA marcado. Estas sondas se desnaturalizaron durante 10 minutos a 80°C y se añadieron sobre las preparaciones. Tras colocar el cubreobjeto, las preparaciones se sellaron con pegamento imedio y se incubaron en una cámara húmeda durante 16 horas a 37°C, excepto cuando la sonda empleada era 10bp, en cuyo caso se incubaron a 30°C. En estos casos la temperatura de incubación se redujo debido al alto contenido en AT de la sonda.

3.2.13.4. Inmunodetección de las sondas

Tras la hibridación las preparaciones se sometieron a los siguientes lavados: tres lavados de 5 minutos en 50% formamida y 2x SSC a 42°C, otros tres lavados de 5 minutos en 0,1x SSC a 60°C. A continuación se incubaron con solución de bloqueo (3% BSA, 4x SSC y 0.1% Tween 20) a 37°C durante 30 minutos en una cámara húmeda oscura. Para detectar las sondas, se usaron los anticuerpos anti-Digoxigenina conjugado con fluoresceína o anti-Digoxigenina conjugado con rodamina (Roche) diluidos hasta una concentración de 2 µg/ml en 1% BSA, 4x SSC y 0.1% Tween 20. En ambos casos la incubación se llevó a cabo durante 30 minutos a 37°C en una cámara húmeda oscura. Posteriormente se realizaron tres lavados de 5 minutos en 4x SSC y 0,1% Tween20 a 42°C. El DNA se tiñó con una solución de DAPI (4',6'-diamidino-fenolindol) a 50 ng/ml en 2x SSC durante 5 minutos y las preparaciones se montaron con una solución "antifading" (Vectashield) y se sellaron con laca de uñas, conservándose a -20°C hasta su observación en un microscopio invertido Axiovert200 (Zeiss) acoplado a una cámara CCD SPOT RT Slider (Diagnostic). Las imágenes fueron obtenidas mediante el programa Metavue (Zeiss) y el procesamiento digital posterior de las mismas se realizó usando el programa Adobe Photoshop.

3.2.14. Inmunofluorescencia e hibridación *in situ* en fibras de cromatina

3.2.14.1. Preparación de fibras de cromatina

Se centrifugaron 5×10^4 células S2 durante 2 minutos a 3500 rpm. El sedimento celular se resuspendió en citrato sódico al 0,5%, y se incubó a temperatura ambiente durante 10 minutos, tras los cuales se centrifugó en una citocentrifuga 4 minutos a 800 rpm, de modo que las células quedaron adheridas a los portaobjetos poli-lisinados. Se sumergieron las preparaciones durante 25 minutos a temperatura ambiente en una solución de lisis, compuesta por 25 mM Tris, 0,5 M cloruro sódico y 1% Triton X-100. A continuación se retiró la solución a una velocidad constante, mediante una bomba peristáltica, de modo que la tensión superficial de la solución, al bajar el nivel de la misma en la cubeta, estirase las fibras de cromatina sobre la superficie del portaobjetos. Se incubaron las preparaciones 15 minutos a

temperatura ambiente en 1x PBST. De nuevo se incubaron en la misma solución de lisis, esta vez durante 15 minutos, tras los cuales se volvió a retirar la solución empleando la bomba peristáltica. Se fijaron las fibras durante 10 minutos en una solución de 1x PBST que contenía formaldehído al 4%, tras lo cual, se lavaron durante 5 minutos en 1x PBS y se procedió a realizar la inmunofluorescencia.

3.2.14.2. Inmunofluorescencia

En el desarrollo de esta tesis doctoral únicamente se empleó un anticuerpo, que reconoce la proteína centromérica CID (homóloga en *Drosophila* de la proteína CENP-A humana). El anticuerpo primario empleado fue anti-CID, de pollo. El anticuerpo se utilizó a una dilución de 1:100, en suero de cabra (Zymed Laboratorios Inc.). El anticuerpo secundario empleado fue anti-pollo IgG conjugado con Alexa 488 (Molecular probes) y se empleó una dilución de 1:500, en el mismo suero de cabra.

Tras bloquear las fibras durante 30 minutos en suero de cabra, la incubación con el anticuerpo primario (anti-CID) se realizó durante toda la noche a 4°C en una cámara húmeda. A continuación, se hicieron tres lavados en PBST (1x PBS y 0.1% Triton) de 5 minutos cada uno a temperatura ambiente, tras los cuales se procedió a la incubación con el anticuerpo secundario (anti-pollo Alexa 488) durante 1 hora a 37°C en cámara húmeda. Las preparaciones se lavaron tres veces durante 5 minutos con PBST y tres veces con PBS.

3.2.14.3. Hibridación *in situ*

El marcaje de las sondas, las condiciones de hibridación y la inmunodetección de las mismas se realizó como se detalla en los apartados 3.2.13.2, 3.2.13.3, y 3.2.13.4. El DNA de las fibras se tiñó con DAPI y las preparaciones se montaron y analizaron tal y como está descrito en el apartado 3.2.13.4.

Resultados

4. RESULTADOS

4.1. El centrómero del cromosoma Y de *Drosophila melanogaster*

4.1.1. Análisis de la región centromérica h18

4.1.1.1. Desarrollo de una estrategia de secuenciación para DNA heterocromático

Aunque los cósmidos, los BACs y los clones de genotecas P1 pueden secuenciarse eficientemente mediante la técnica “shotgun” (Wilson y Mardis, 1997), en muchas ocasiones, cuando el fragmento genómico inserto en el clon es DNA repetido, esta técnica de secuenciación no funciona. Las razones son varias: dificultad a la hora de subclonar, eliminación de ciertas secuencias repetidas, reordenamientos, además del gran problema que conlleva el ensamblado de las lecturas obtenidas. En la actualidad grupos especializados en “finalización de la secuencia” están realizando grandes esfuerzos para obtener la secuencia completa de clones heterocromáticos. Para resolver este problema, en esta tesis doctoral se ha modificado la técnica de secuenciación mediante transposones para permitir secuenciar satisfactoriamente, y de manera automatizable, cualquier clon heterocromático de gran tamaño (por ejemplo BACs), independientemente de la naturaleza y composición de su secuencia.

4.1.1.2. Obtención del transposón artificial Tn599

La técnica de secuenciación se basa en el empleo de un transposón del tipo Tn5 para generar, *in vitro*, una genoteca de clones con el transposón insertado aleatoriamente a lo largo de todo el fragmento de interés. Los BACs poseen un origen de replicación de bajo número de copias y por ello son el vector idóneo para trabajar con DNA repetido altamente inestable. Sin embargo, a la hora de manejar cultivos bacterianos y de extraer el DNA es preferible manejar volúmenes pequeños con mayor número de copias por célula. Por lo tanto, en la construcción del transposón se ha partido del plásmido pMOD2<oriV/KAN-2>, que suministra el origen de replicación inducible oriV. Tras la inducción de este origen de replicación, mediante la producción de la proteína de replicación TrfA, se generan múltiples copias del clon en la célula huésped. Esta

capacidad sólo se utiliza en el paso previo a la extracción del DNA, disminuyendo la posibilidad de que se produzcan modificaciones en la secuencia de interés. Además, en el transposón se han insertado las dianas de reconocimiento para las enzimas *PI-SceI* e *I-SceI*. Estas dianas son indispensables para poder mapear el sitio de inserción del transposón en cada clon. Posteriormente, los datos del mapeo se combinarán con las lecturas de secuencia para conseguir un ensamblado adecuado. Al incluir la información de la posición nunca se ensamblarán erróneamente secuencias similares que provengan de distintos lugares del DNA repetido.

Tanto *PI-SceI* como *I-SceI* son endonucleasas tipo “homing”, esto es, DNAsas de cadena doble que reconocen una diana asimétrica muy larga; 18 bp para *I-SceI* y 39 bp para *PI-SceI*. Por esto, las dianas de las endonucleasas del tipo “homing” son poco frecuentes. Por ejemplo, se ha estimado que una diana de reconocimiento de 18 bp de longitud ocurre una vez por cada 7×10^{10} bp de secuencia. Esto equivale a decir que la secuencia diana se encontrará una sola vez en 20 genomas de mamíferos (Jasin, 1996). Sin embargo, al contrario de lo que ocurre con las enzimas de restricción habituales, las endonucleasas “homing” toleran cierta degeneración en la secuencia diana de reconocimiento (Argast et al., 1998; Gimble y Wang, 1996). A pesar de esto último, la probabilidad de encontrar estas dianas de reconocimiento en una secuencia de interés es casi nula. Así, mediante una simple digestión, se podrá medir el sitio de inserción del transposón modificado con estas dianas, siempre que el vector empleado en la construcción de las genotecas lleve al menos una de ellas. En la actualidad, la mayoría de las genotecas incluye una de estas dianas (<http://bacpac.chori.org/vectorsdet.htm>).

Teniendo en cuenta todas estas consideraciones, la construcción del plásmido fue llevada a cabo del modo que se describe a continuación. Se partió del plásmido pMOD2<oriV/KAN-2> (cedido por Epicentre) el cual fue digerido con *HindIII*, se hicieron extremos romos y se defosforiló. El vector fue entonces ligado a un fragmento de DNA de 600 bp que contenía la diana de reconocimiento para *I-SceI*. Este fragmento fue obtenido, a su vez, mediante digestión doble con las enzimas *EcoRI* y *Sall* del plásmido pSCM522

(Monteilhet et al., 1990), y también se hicieron romos sus extremos, para hacer posible la ligación con el vector. Una vez se obtuvo el plásmido pMOD2<oriV/KAN-2/I-SceI>, éste fue digerido con la enzima *SacI*, se hicieron extremos romos, se deforforiló y se ligó al fragmento de 750 bp que incluía la diana para la enzima *PI-SceI*. Este fragmento se había obtenido previamente mediante digestión doble con las enzimas *HindIII* y *PstI*, a partir del plásmido pSBVDEX (Gimble y Thorner, 1993) y se habían hecho sus extremos romos. La construcción final obtenida se trata del plásmido pMOD2<oriV/KAN-2/I-SceI/PI-SceI>, del cual se obtiene el transposón Tn599 por digestión doble con *PshAI* y con *ApaLI*. El tamaño del transposón es de 3,3 kb. La figura 7 muestra la construcción del plásmido pMOD2<oriV/KAN-2/I-SceI/PI-SceI> y una representación esquemática del transposón Tn599.

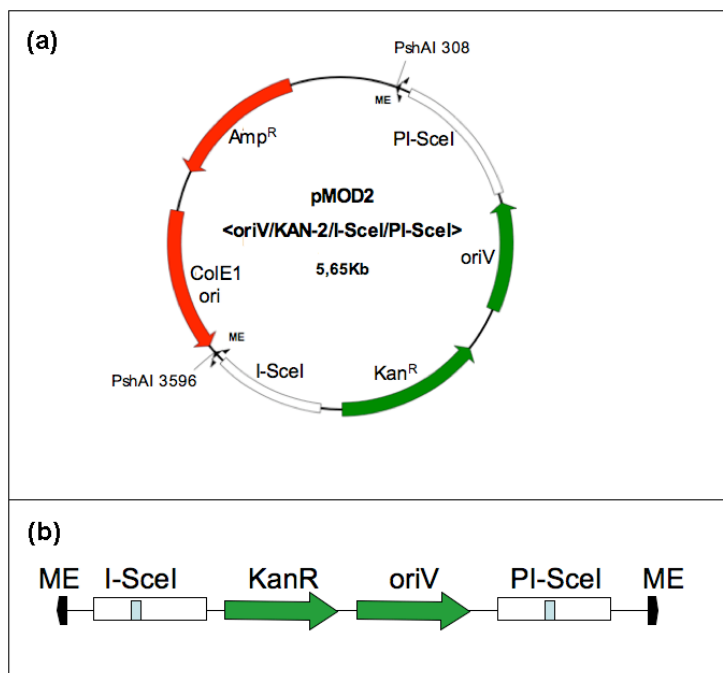


Figura 7. Plásmido pMOD2<oriV/KAN-2/I-SceI/PI-SceI> y transposón Tn599.

(a) Esquema del plásmido pMOD2<oriV/KAN-2/I-SceI/PI-SceI>. (b) Esquema del transposón Tn599.

Resumiendo, el transposón contiene los siguientes rasgos característicos: (i) dos secuencias “mosaico final “ (ME, del inglés “mosaic end”) en los extremos del transposón, necesarios para que el transposón se integre, (Zhou et al., 1998), (ii) secuencias conocidas adyacentes a los ME que sirven como lugares

de unión de los cebadores que serán empleados para secuenciar la región del inserto contigua al transposón, (iii) un gen de resistencia a kanamicina, que permitirá seleccionar aquellos clones en los que se haya insertado el transposón, (iv) un origen de replicación inducible oriV, que transformará el BAC de “copia única” a “múltiples copias”, (v) y una diana de reconocimiento para la endonucleasa *I-SceI* y otra para *PI-SceI* (ver Figura 7.b).

4.1.1.3. Validación de la nueva técnica de secuenciación

Para evaluar la estrategia, se decidió probarla en un BAC heterocromático de la genoteca RPCI-98 de *D. melanogaster*, construida para el proyecto de secuenciación del genoma de *Drosophila* llevado a cabo en Berkeley (Berkeley Drosophila Genome Project, BDGP). El vector de clonación utilizado en la construcción de la genoteca fue el vector pBACe3.6, que tiene una diana de reconocimiento para la enzima *PI-SceI* en su posición 11371-11407 (Hoskins et al., 2000). El BAC elegido fue el BACR26J21, pues según los datos previos, estaba compuesto por unas 100 kb de repeticiones de 3,1 kb del satélite 18HT, algunos elementos *HeT-As* truncados y algún transposón, como el elemento *F* y el elemento *copia* (Abad et al., 2004a; Agudo et al., 1999; Losada et al., 1997). Con objeto de saber si el nuevo abordaje de secuenciación mejoraba las técnicas estándar existentes, el BACR26J21 se secuenció independientemente en dos laboratorios. La secuenciación y ensamblado con la nueva estrategia fue realizada en el CBMSO, mientras que la secuenciación mediante los protocolos estándar fue realizada por el grupo de “finalización de la secuencia” del Wellcome Trust Sanger Institute (WTSI), bajo la supervisión de la Dra. Whitehead.

4.1.1.4. Obtención de una genoteca del BACR26J21 mediante transposición del transposón Tn599

El primer paso consistió en la formación de complejos de transposición incubando 1 µg del transposón Tn599 (fragmento de 3,3 kb) con 10 µg/µl de transposasa hiperactiva durante 2 horas a 37°C en 40 µl de un tampón que contiene 50 mM Tris-acetato (pH 7.5), 150 mM acetato potásico, 10 mM acetato magnésico y 4 mM espermidina. Posteriormente se incubaron 14 µl de la reacción de formación de estos complejos con 1 µg de DNA del BACR26J21

durante 2 horas a 37°C en un volumen final de 20 μ l. Las reacciones se pararon con 0,1% SDS y 10 minutos a 70°C. Tras desalación en conos de agarosa, se transformó y se seleccionó con 30 μ g/ml kanamicina y 12,5 μ g/ml cloranfenicol.

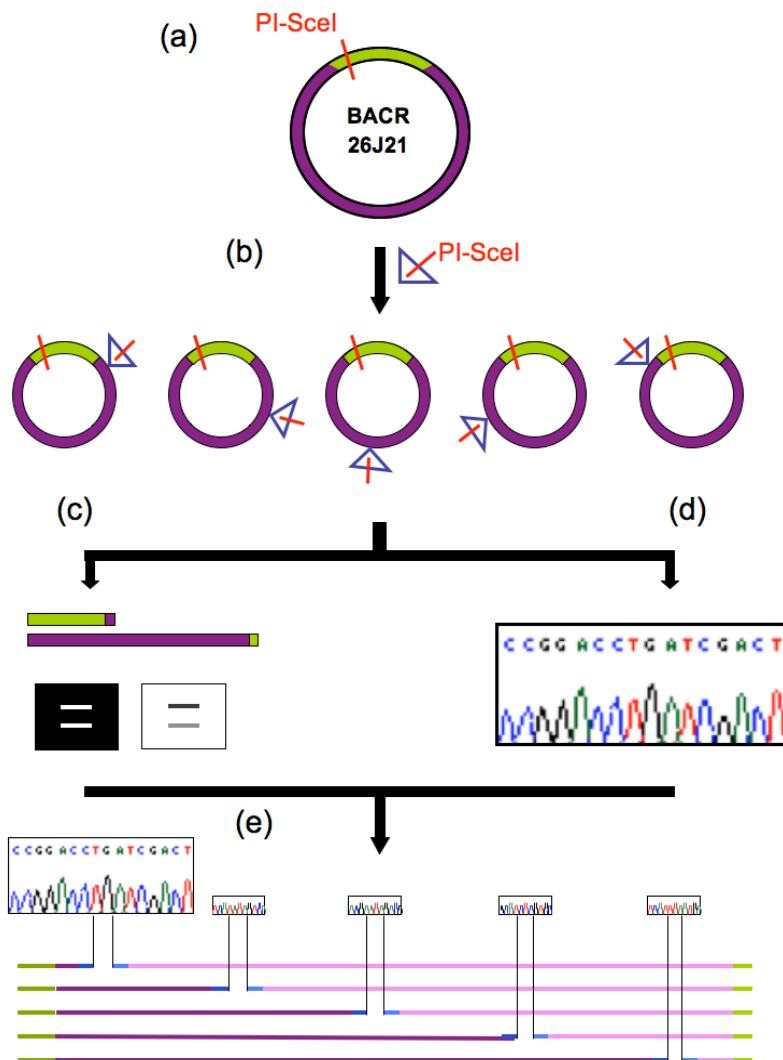


Figura 8. Esquema de la técnica de secuenciación mediante trasposición, incorporando la información posicional.

(a) BACR26J21, (b) transposición *in vitro* para generar una genoteca, (c) mapeo de la posición de inserción del transposón, (d) secuenciación y (e) ensamblado, empujando la posición.

En un único evento de transposición *in-vitro* del transposón Tn599 sobre el BACR26J21 se obtuvieron alrededor de 2.500 transformantes, generándose así una genoteca del BACR26J21 con el transposón Tn599 insertado al azar. (ver Figura 8, a y b).

4.1.1.5. Mapeo de la inserción del transposón Tn599.

Es importante mencionar que antes de comenzar se realizó una estimación del número de clones que convendría procesar. La longitud del fragmento heterocromático presente en el BACR26J21 se midió mediante electroforesis de campo pulsado, tras la digestión del BAC con la enzima de restricción *NotI* (datos no mostrados) y se estimó en 160 kb. Además se consideró que era necesario obtener una cobertura de lectura de entre 4x y 5x, esto es, que cada base de la secuencia final hubiese sido secuenciada en 4-5 lecturas distintas, como media. Para realizar este cálculo se consideró que 700 nucleótidos es la longitud media de secuencia de calidad que se obtiene rutinariamente en un secuenciador de capilares ABI 3730xl. Con estos datos se procedió al cálculo de la siguiente manera:

$$\text{Nº clones} = \frac{\text{Nº nucleótidos totales} \times \text{cobertura de lectura}}{2 \text{ lecturas/clon} \times 700 \text{ nucleótidos/lectura}}$$

De este modo se obtuvieron los valores que muestra la tabla 2 para el número de clones a procesar si se quiere obtener una cobertura de lectura de 4x y de 5x.

Tabla 2. Relación entre el número de clones y la cobertura de lectura de la secuencia.

Cobertura de lectura	Nº de clones
4x	457
5x	571

Tomando esos valores obtenidos como el número de clones a procesar, y asumiendo que habría un pequeño porcentaje de clones que podrían fallar (por inserción de más de un transposón, por ejemplo), se decidió medir el sitio de inserción del transposón en 620 clones, para tener un amplio margen.

Se inocularon cultivos de los 620 clones BACR6J21-Tn599 en LB con kanamicina y cloranfenicol. Cuando la absorbancia de dichos cultivos (medida a una longitud de onda de 590 nm) era de 0.2–0.3, se añadió L-arabinosa

hasta una concentración final del 0.01% (para inducir la expresión del gen *TrfA*) y los cultivos se continuaron creciendo durante 4–5 horas más, antes de centrifugar los mismos para extraer el DNA.

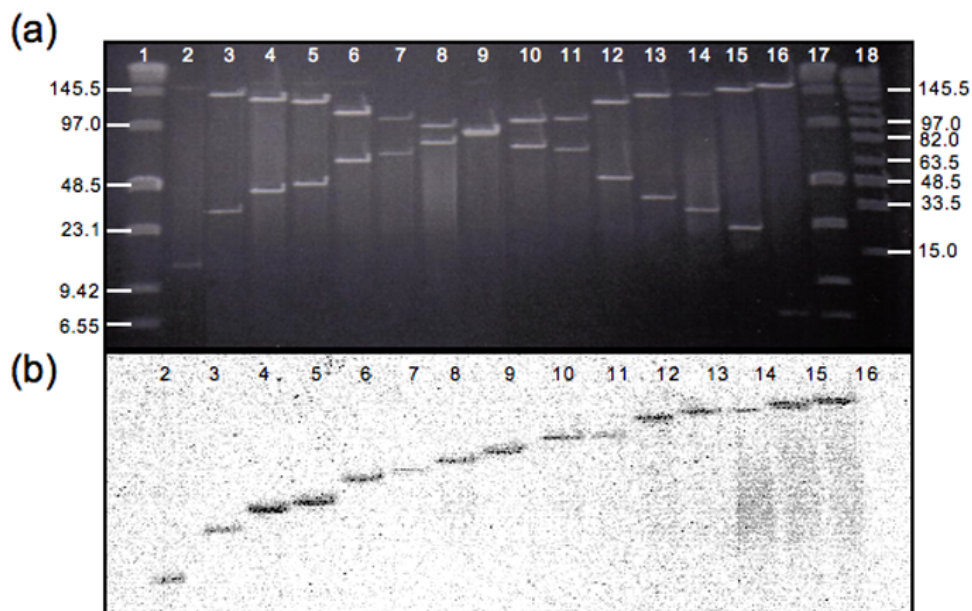


Figura 9. Mapeo de la inserción del transposón Tn599 en 15 clones.

(a) Imagen de un gel con el DNA digerido con *PI-SceI* y fragmentado mediante electroforesis de campo pulsado (1% agarosa, 6 V/cm, pulso de 0,4 -10 seg, durante 20 horas a 14°C). (b) Hibridación del DNA transferido, con la sonda P.

A continuación, se procedió a mapear el sitio de inserción del transposón en cada clon, para lo cual se digirió DNA extraído de los 620 clones con la enzima de restricción *PI-SceI* (figura 8c). En cada clon sólo existen dos sitios de reconocimiento; uno en el vector pBACe3.6 y el otro en el transposón Tn599. Tras la digestión de los clones, los dos fragmentos generados se separaron mediante electroforesis en campo pulsado y con ayuda de los marcadores de peso molecular se estimó el tamaño del fragmento más pequeño. Únicamente se midió la banda inferior porque en las condiciones de electroforesis empleadas el rango de máxima separación abarca hasta unas 90 kb, perdiéndose resolución en tamaños superiores. El tamaño de la banda superior se calculó restando la inferior del tamaño total del clon. Sin embargo, si bien se obtuvieron los tamaños de ambos fragmentos, todavía no se tenía suficiente

información para asignar la posición del transposón en el BAC, puesto que existen dos posibilidades para cada par de medidas: las dos posiciones serían simétricas si se considera la diana de *PI-SceI* del vector como punto de origen de la medición. Por este motivo, el DNA del gel fue transferido a una membrana de nylon cargada positivamente y ésta fue hibridada con la sonda P marcada radioactivamente (figura 8c y figura 9). La sonda P deriva de una región del vector pBACe3.6, y se obtuvo mediante PCR utilizando el vector pBACe3.6 como molde de DNA y los cebadores P-Fw y P-Rv. La secuencia de los cebadores y las condiciones de la reacción de PCR están detallados en la tabla 3.

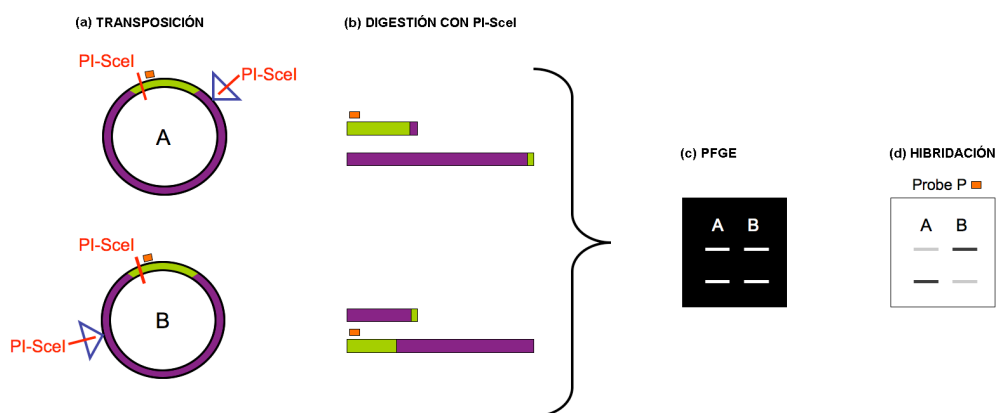


Figura 10. Mapeo de la inserción del transposón en dos clones en los que el transposón se ha insertado equidistantemente respecto a la diana *PI-SceI* del vector.

(a) Clones A y B, (b) Digestión con *PI-SceI*, (c) PFGE y (d) Hibridación con la sonda P.

La figura 10 muestra un ejemplo de dos clones (denominados A y B) con el transposón insertado en las posiciones simétricas respecto de la diana *PI-SceI* del vector. Ambos clones, tras la digestión con *PI-SceI* generan una pareja de fragmentos de tamaños semejantes. Tomando únicamente las medidas de los fragmentos, no se puede distinguir la posición relativa de cada uno de ellos (ver figura 10a. A y B). Sólo después de la hibridación con la sonda P se pueden discernir los dos fragmentos, y, por lo tanto, situarlos en su posición correcta a lo largo del BAC (figura 10. A y B). En la mayoría de los clones analizados la digestión con la enzima *PI-SceI* generó dos fragmentos de DNA, pero en ocasiones se generaron tres o más fragmentos, probablemente, como

consecuencia de una inserción múltiple del transposón en una molécula del BAC. Estos clones fueron descartados.

Tabla 3. Cebadores y condiciones de la PCR para obtener la sonda P.

<u>Cebadores</u>			
P-Fw: 5'AAGGCCGTAATATCCAGCTG 3'			
P-Rv: 5'CTTCGTGTAGACTTCCGTTG 3'			
<u>Programa de PCR</u>			
Desnaturalización inicial	94°C, 5 min		
Desnaturalización	94°C, 1 min	}	30 ciclos
Anillamiento	55°C, 30 seg		
Extensión	72°C, 1 min		
Extensión final	72°C, 10 min		

Por otra parte, también se estudió la posibilidad de obtener la orientación de inserción del transposón en cada clon, que supondría más información, y ayudaría en el proceso de ensamblaje. Para esto, se diseñó otra sonda, la sonda T, que hibrida en el transposón a un lado de la diana de reconocimiento de *PI-SceI*. Esta sonda se obtuvo mediante PCR utilizando el transposón Tn599 como molde de DNA y los cebadores T-Fw y T-Rv. La secuencia de los cebadores y las condiciones de la reacción de PCR están detallados en la tabla 4. El producto de PCR obtenido fue marcado radioactivamente. Tras la hibridación con la sonda P, las membranas de nylon con el DNA digerido fueron lavadas a alta temperatura para levantar los restos de la sonda P e hibridados de nuevo con la sonda T. Como esta segunda hibridación resultaba muy laboriosa y dado que la información que proporcionaba no era tan crítica (pues el propio proceso de ensamblaje hace que se orienten las lecturas entre sí), se decidió usar solamente cuando fuese necesario resolver regiones que presentasen complicaciones.

Tabla 4. Cebadores y condiciones de la PCR para obtener la sonda T.

Cebadores

T-Fw: 5' CACGGTTGATGAGAGCTTTG 3'

T-Rv: 5' TGGCCTGTTGAACAAGTCTG 3'

Programa de PCR

Desnaturalización inicial	94°C, 5 min	} 30 ciclos
Desnaturalización	94°C, 1 min	
Anillamiento	55°C, 30 seg	
Extensión	72°C, 1 min	
Extensión final	72°C, 10 min	

4.1.1.6. Comprobación de la aleatoriedad de inserción del transposón en regiones de DNA altamente repetido

Kang y colaboradores mostraron que el transposón Tn5 reconoce preferentemente una secuencia asimétrica muy degenerada y que la flexibilidad de esta diana es una faceta crítica en la aleatoriedad observada en la inserción del transposón Tn5 (Kang et al., 2004). Sin embargo, se consideró necesario comprobar la aleatoriedad en la inserción del transposón Tn599, en el inserto heterocromático del BACR26J21. Como se observa en la figura 11, la inserción tuvo lugar aleatoriamente.

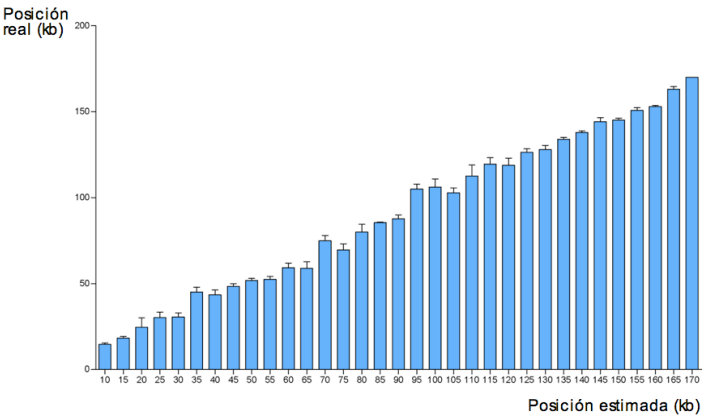


Figura 11. Aleatoriedad en la inserción del transposón Tn599 a lo largo del BACR26J21.

4.1.1.7. Secuenciación de los clones

El siguiente paso de la estrategia fue secuenciar el DNA adyacente al transposón (ver Figura 8.d). En la secuenciación se usaron dos cebadores, TN-FP y TN-RP (ver tabla 5), que hibridan con el DNA en los extremos del transposón (regiones contiguas a los ME) y permiten la lectura en dirección al DNA genómico. Los primeros nucleótidos presentes en las lecturas obtenidas corresponden a la secuencia ME. Cuando el transposón Tn599 se inserta genera una duplicación de 9 bp. De este modo, el transposón queda flanqueado por estas 9 bp. Este comportamiento sirve como una herramienta muy útil en el ensamblado posterior. Como consecuencia, cada lectura comienza con secuencia del ME, seguida por 9 pb correspondientes a la secuencia de interés, que habrán sido duplicadas y la secuencia continúa tanto como permitan las tecnologías empleadas. En este análisis, utilizando secuenciadores ABI 3730xl y la química de terminadores marcados “Big Dye Terminators v3.1” de Applied Biosystems, se obtuvo una longitud media de secuencia de calidad de 800 nucleótidos por lectura.

Tabla 5. Cebadores y condiciones de la reacción de secuenciación.

<u>Cebadores</u>		
TN-FP: 5' GCCAACGACTACGCACTAGCCAAC 3'		
TN-RP: 5' GAGCCAATATGCGAGAACACCCGAGAA 3'		
<u>Condiciones de la reacción de secuenciación</u>		
Desnaturalización inicial	95°C, 5 min	} 35 ciclos
Desnaturalización	95°C, 30 sec	
Anillamiento	55°C, 20 seg	
Extensión	60°C, 4 min	
Extensión final	60°C, 5 min	

4.1.1.8. Ensamblado de la secuencia del BACR26J21

El ensamblado de la secuencia constituye el último paso de la estrategia. Es el paso en que se une toda la información recopilada para solapar las lecturas de secuencia, teniendo en cuenta la posición que ocupa cada lectura a lo largo del BAC (Figura 8.e). Los programas PreGap4 y Gap4 fueron utilizados para

ensamblar la secuencia. Como se ha mencionado anteriormente, cuando el transposón se inserta se produce una duplicación de 9 pb, que están presentes hacia el comienzo de las dos lecturas obtenidas para cada clon. Tras la eliminación programada de la secuencia correspondiente al ME del transposón, cada pareja de lecturas de un mismo clon fueron unidas mediante alineamiento de la región duplicada de 9 pb, creando “minicontigs” para cada clon de, aproximadamente, 1,5 kb. Esta unión es muy importante cuando se trata con secuencias altamente repetidas (con una identidad de secuencia superior al 99% entre distintas unidades), pues cuanto más larga sea una secuencia más fácil será encontrar diferencias de hasta una sola base, disminuyéndose así la posibilidad de errores durante el ensamblaje. A continuación, por cada 20 kb del BAC se generó un subproyecto (ver la distribución de los ocho subproyectos en la tabla 6). En cada subproyecto se incluyeron los “minicontigs” de los clones que, según la posición de inserción del transposón, pertenecían a ese rango.

Tabla 6. Rango de los subproyectos generados en el ensamblaje de la secuencia del BACR26J21.

Subproyecto	Intervalo de secuencia
1	0- 20 kb
2	20- 40 kb
3	40- 60 kb
4	60- 80 kb
5	80- 100 kb
6	100- 120 kb
7	120- 140 kb
8	140 kb- final

Posteriormente, los “minicontigs” de cada subproyecto se ensamblaron por solapamiento de secuencia, tomando siempre en consideración la posición relativa de cada clon dentro del subproyecto. El tamaño de 20 kb para cada subproyecto se decidió teniendo en cuenta que la medida del tamaño de los fragmentos separados mediante electroforesis de campo pulsado está asociada a un error inevitable, estimado en un 10%. Como ya se ha comentado, este error se trató de minimizar midiendo únicamente la banda de

menor tamaño. Cuando no se encontraron más solapamientos dentro de un subproyecto, éste fue unido al subproyecto contiguo para continuar el ensamblaje. Esto se repitió sucesivamente hasta que todas las lecturas fueron ensambladas en un único “contig” (ver figura 12) y la secuencia completa para el BACR26J21 se depositó en la base de datos del Genbank, con número de acceso CU076040.

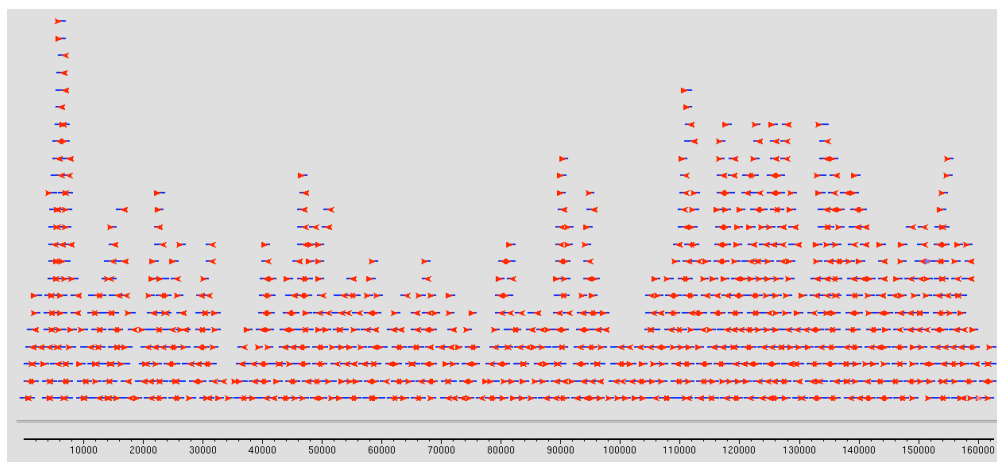


Figura 12. Distribución de lecturas en el ensamblado de la secuencia del BACR26J21.

4.1.1.9. Comparación de las secuencias obtenidas para el BACR26J21 en el CBMSO y el WTSI

Para evaluar la eficacia de la nueva técnica de secuenciación se estableció una colaboración con Jane Rogers, del The Wellcome Trust Sanger Institute (WTSI), de modo que el mismo BAC fuese secuenciado por un grupo especializado en “finalización de la secuencia” de genomas. La estrategia empleada en el WTSI para secuenciar el BACR26J21 consistió, en primer lugar, en la generación de una genoteca de subclones de 4-6 kb mediante la técnica “shotgun”. Las lecturas obtenidas, utilizando cebadores del vector M13, fueron ensambladas usando el programa Phrap (desarrollado por P. Green). La confirmación de la integridad del ensamblado se realizó mediante los programas Gap4 (Staden, 1996), Orchid (desarrollado por D. Flowers) y Confirm (desarrollado por T. K. Attwood). Cuando se observaron errores de ensamblaje, las copias fueron comparadas entre sí para identificar diferencias en una sola base. Estas diferencias fueron marcadas y usadas, posteriormente,

para reconocer regiones de secuencia única dentro del satélite de DNA. Dado que este abordaje resultó ser insuficiente, se procedió a emplear la estrategia “LIL-TIL”. Esta estrategia consiste, inicialmente, en la generación de una genoteca con un inserto de gran tamaño, entre 9 y 12 kb, (en inglés “Large Insert Library”, LIL). Una vez secuenciados los extremos de cada inserto, esta información se combina con la información del ensamblado anterior, para seleccionar algunos de los clones de la genoteca LIL. A continuación, los clones seleccionados son secuenciados mediante transposición, para lo cual se construyen genotecas generadas por transposición (en inglés “Transposon Insertion Libraries”, TIL). La secuencia de estos clones fue combinada con la información del ensamblado general previo, con información de análisis de restricción del BAC y con la información existente de las parejas de lectura del “shotgun” para proseguir el ensamblaje del BACR26J21, hasta finalizar con un único “contig”.

Las dos secuencias obtenidas para el BACR26J21, fueron comparadas mediante el programa BLAST2, y se observó que eran casi idénticas. Como era de esperar, la región de 68 kb que expande desde el satélite 18HT hasta el extremo opuesto del inserto del BAC era idéntica en ambas versiones, mientras que la única diferencia encontrada se hallaba en la parte del satélite de DNA del BAC. La versión obtenida en el WTSI tenía una unidad de repetición del satélite menos que la versión del CBMSO. Dado que el análisis de restricción del clon no proporciona información suficiente para discernir entre ambas versiones, y a falta de otros criterios y técnicas sensibles que determinen cuál de las dos versiones es la correcta, se decidió que la secuencia del BACR26J21 obtenida en el CBMSO era más fiable, dado que en su ensamblado se tuvo en cuenta la información de posición.

4.1.1.10. Organización de la región centromérica de h18

La secuencia ensamblada del BACR26J21 fue analizada mediante el programa BLAST contra las bases de datos de FlyBase y mediante análisis de “Dotplot”. La secuencia ha sido anotada manualmente (Anexo I), y dicha anotación ha sido utilizada para hacer una reconstrucción de los procesos evolutivos que ha experimentado la región h18.

El análisis de la secuencia muestra, como ya se había anticipado, que alrededor de dos tercios del inserto de este BAC corresponden al satélite 18HT, y que el tercio restante estaba formado por retrotransposones *HeT-A* degenerados. A lo largo de toda la región existen también inserciones de transposones y pequeñas duplicaciones segmentales (Figura 13.a).

Estudiando en profundidad la transición entre el satélite 18HT y los *HeT-A* degenerados, se pudo establecer con exactitud cual fue el fragmento de DNA que sufrió la amplificación masiva, generando un satélite con una unidad de repetición de ~3,1 kb. En el extremo del inserto correspondiente al satélite se observó que la unidad de repetición ya no era de ~3,1 kb, sino de ~2 kb, hecho que había sido previamente detectado (Abad et al., 2004a; Agudo et al., 1999). Esta unidad más pequeña fue generada, probablemente, mediante una delección y, posteriormente, fue amplificada. Asimismo se encontraron otras delecciones en las unidades del satélite, siendo alguna de gran tamaño y múltiples pequeñas.

A lo largo del satélite se observó la presencia de tres transposones: *1360*, *412* y *copia*. A continuación del satélite hay *HeT-As* degenerados, entremezclados con los elementos transponibles *mdg1*, *diver*, *1731* y *F*. Como el elemento *1731* no muestra degeneración, es de suponer que su inserción ha sucedido recientemente. También se encontró una duplicación segmental de la región eucromática 42A, presente en dos posiciones distintas de la secuencia del BACR26J21. A primera vista se observó que, además de la inserción de transposones, habían ocurrido amplificaciones en distintas partes de la región heterocromática presente en el BAC. Por este motivo se decidió analizar la secuencia con mayor detalle, para hacer una reconstrucción de los procesos evolutivos que podrían haber ocurrido. Este análisis permitiría verificar si, efectivamente, la región de origen de h18 pudo tratarse de un telómero.

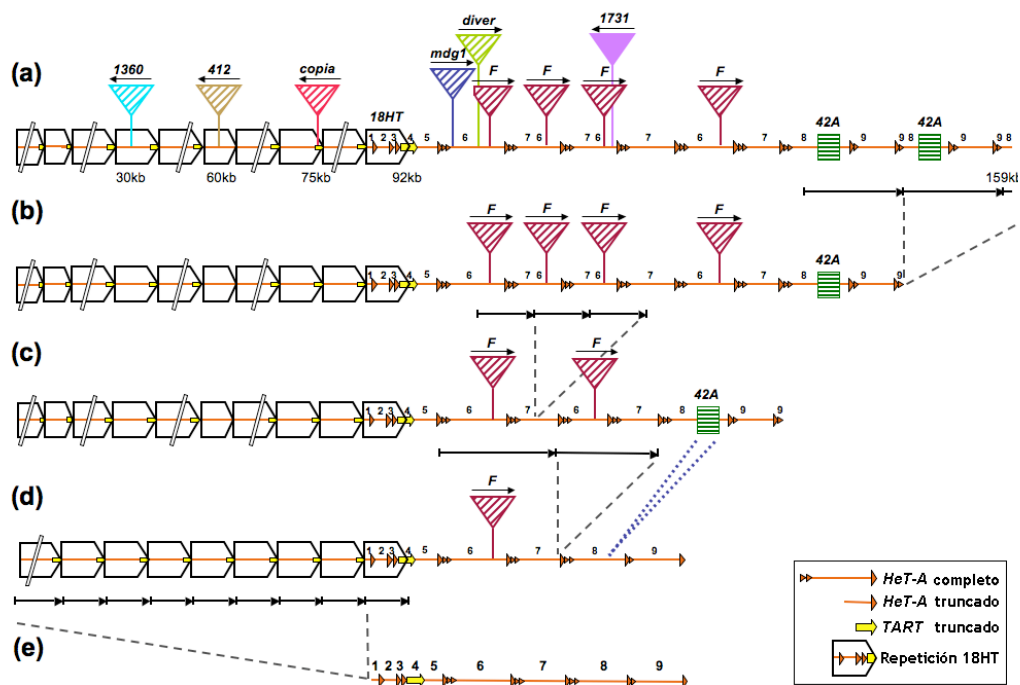


Figura 13. Deconstrucción de la secuencia del BACR26J21

(a) Representación esquemática de la secuencia de la región centromérica h18 del cromosoma Y. Los triángulos coloreados representan transposones, y la caja verde, una duplicación segmental. (b), (c) y (d) Regreso al pasado mediante "Paleontología cromosómica". Las amplificaciones se representan con líneas discontinuas. (e) Representación esquemática de la región telomérica ancestral. El número de acceso en el Genbank para la secuencia del BACR26J21 es CU076040.

4.1.1.11. La región centromérica h18 deriva de un telómero

Para la deconstrucción de la región h18 se comenzó eliminando los transposones que se hallaban presentes una vez en la secuencia, asumiendo que eran sucesos que habían ocurrido de manera aislada durante la evolución. Basándose en la homología de secuencia encontrada para las distintas zonas, se propone una sucesión de eventos de amplificación que podrían haber sucedido durante la evolución de esta región del cromosoma (ver figura 13.a-e). De acuerdo con este análisis, los cuatro elementos *F* derivarían de una única transposición y, posteriormente, la región en la cual se insertó habría sido amplificada varias veces. Esto mismo habría sucedido en la región que contiene la duplicación segmental de 42A, salvo que en este caso la amplificación parece haber ocurrido en una sola ocasión.

Tras haber "limpiado" la región de h18 presente en el BACR26J21 y haberla

simplificado a su estructura ancestral más probable, mediante lo que podría llamarse “paleontología cromosómica”, se obtuvo una sucesión de nueve elementos teloméricos orientados en disposición cabeza-a-cola. Entre estos nueve elementos hay cuatro elementos *HeT-A* truncados (los elementos 1, 2, 3 y 5 en la figura 13.e), un elemento *TART* truncado (el elemento 4) y cuatro elementos *HeT-A* enteros (los elementos 6, 7, 8 y 9). Los *HeT-A* 1, 2 y 3, y parte del elemento *TART* conforman la unidad de repetición de 3,1 kb que originalmente se amplificó para dar lugar al satélite 18HT, previamente descrita (Agudo et al., 1999; Danilevskaya et al., 1993). Se pudo confirmar que los elementos *HeT-A* 6, 7, 8 y 9 estaban enteros porque se detectaron secuencias de la ORF, secuencias de las regiones 3' UTR y 5' UTR e incluso, se pudo comprobar que las regiones 5' UTR presentaban los motivos característicos de 8-17 bp que derivan de la región 3' del elemento *HeT-A* anterior al *HeT-A* completo a partir del cual se retrotranscribe cada elemento *HeT-A* (Danilevskaya et al., 1997). De hecho, es precisamente el número y la composición exacta de estos motivos en el extremo 5' UTR lo que ha sido usado para distinguir los distintos elementos *HeT-A*. Finalmente es importante destacar que la secuencia ancestral que ha sido reconstruida tiene una longitud similar, de unas 30 kb, a la de las series de retroelementos teloméricos presentes en los telómeros de *Drosophila* en la actualidad (Abad et al., 2004b; George et al., 2006). Dado que la presencia de retroelementos teloméricos enteros es una característica típica de los telómeros de *D. melanogaster* (Abad et al., 2004b; George et al., 2006) y que la orientación de éstos y su longitud total son también las adecuadas, se ha confirmado que la región centromérica h18 derivó de un telómero.

4.1.2. Análisis de la región centromérica h17

4.1.2.1. Identificación de BACs de la región centromérica h17 del cromosoma Y de *D. melanogaster*

El conocimiento previo de las regiones centromérica y pericentromérica del cromosoma Y de *Drosophila melanogaster* se obtuvo avanzando desde los brazos del cromosoma mediante “walking”, que implica la obtención de regiones de secuencia única y su utilización como sondas en el rastreo de las genotecas de BACs. Una vez se obtienen los BACs que dan positivo en cada

rastreo, se realizan mapas de restricción, y se observa si los BACs solapan (aparición de bandas comunes a varios BACs). Además se confirma la localización de esos BACs al cromosoma de interés mediante la técnica citogenética FISH. Esto se realiza sucesivamente, hasta que se da con una región que carece de otras secuencias únicas que permitan continuar el “walking”. De este modo se consiguió avanzar por el brazo corto, YS, a través de un tándem de elementos *R1*, otro tándem de elementos *Circe*, otro de elementos *HeT-A*, el satélite 18HT, un tándem de elementos *Doc2* y, finalmente, un tándem del pseudogen *Mst77*, como muestra en la figura 14 (Abad et al., 2004a; Agudo et al., 1999; Losada et al., 1997). Desafortunadamente, la ausencia de secuencias únicas en la zona del pseudogen *Mst77*, no permitió continuar más allá del mismo. Por otro lado, el rastreo de una genoteca de BACs con una sonda de *TART* (un retrotransposón telomérico) proporcionó clones de la región h16 (brazo largo YL). Al analizar la secuencia de dichos BACs se observó que había elementos *TART* entremezclados con transposones, pero no se encontró ninguna región de secuencia única con la que poder continuar avanzando hacia h17. Encontrar regiones inabordables mediante esta estrategia es algo común cuando se trabaja en la heterocromatina.

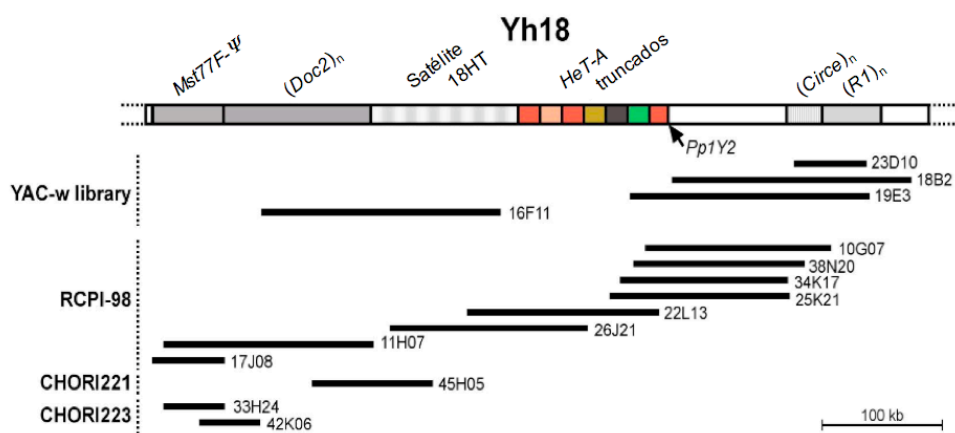


Figura 14. “Contig” de la región h18 del cromosoma Y, construido mediante “walking”.

El Proyecto de Secuenciación de la Heterocromatina de *Drosophila* (DHGP) generó una genoteca de inserciones P en la heterocromatina de *D.*

melanogaster, como herramienta adicional para su análisis. Las inserciones P, además de servir para obtener mutantes, son de gran utilidad en el estudio del genoma, pues proporcionan regiones de copia única que pueden servir como punto de entrada tanto para análisis estructurales como funcionales. En cada inserción se determinaron las secuencias que flanquean el elemento P y se localizó éste mediante hibridación *in situ* a cromosomas mitóticos. Dado el vacío existente en el conocimiento de la región heterocromática h17, y la incapacidad de avanzar mediante “walking”, se decidió buscar entre los estudios realizados con elementos P para ver si algún dato podía referirse a esta región.

Afortunadamente, en 2002 Maggert y Golic publicaron su estudio sobre la impronta genética que afecta al cromosoma Y de *D. melanogaster*. En este trabajo analizaron la expresión y localización de veintitrés inserciones de un elemento P, SUPorP, que contiene dos genes marcadores: *white+* y *yellow+* (Maggert y Golic, 2002). Entre los resultados publicados en ese trabajo se observó que la inserción B783.2 mostraba la máxima represión de ambos genes y localizaba en la región h17-h18. Además, la secuencia adyacente a esta inserción no correspondía a ningún transposón o DNA satélite conocidos. La localización citogenética de este elemento en la región centromérica del cromosoma Y, y, fundamentalmente, la alta represión a la que se encontraba sometido estaría de acuerdo con que la inserción hubiese tenido lugar en una localización muy próxima al centrómero o, incluso, en el centrómero mismo. Por este motivo, se utilizó la secuencia del punto de inserción para buscar secuencias homólogas en las bases de datos de *D. melanogaster*. Estas bases de datos son actualizadas regularmente, a medida que se confirman los nuevos datos de secuencia. Tras esta búsqueda se observó que las secuencias con mayor similitud correspondían a los extremos de dos BACs: BACN06I07 y BACN14H20. Estos BACs pertenecen a la genoteca de *D. melanogaster* preparada a partir de DNA genómico digerido parcialmente con la enzima de restricción *Nde*II, que fue construida en el Centro de Estudios de Polimorfismos Humanos, en Francia. A continuación, se realizaron nuevas búsquedas con la secuencia del otro extremo de cada uno de estos BACs. En el extremo del BACN06I07 se hallaba un transposón *roo*, mientras que la

secuencia del extremo del BACN14H20 tenía alta similitud con los extremos de dos nuevos BACs de la genoteca RPCI-98: BACR07N15 y BACR36F19 (ver Figura 15). El análisis de restricción de estos BACs con las enzimas *EcoRI* y *HindIII* mostró que estos BACs derivaban de la misma región.

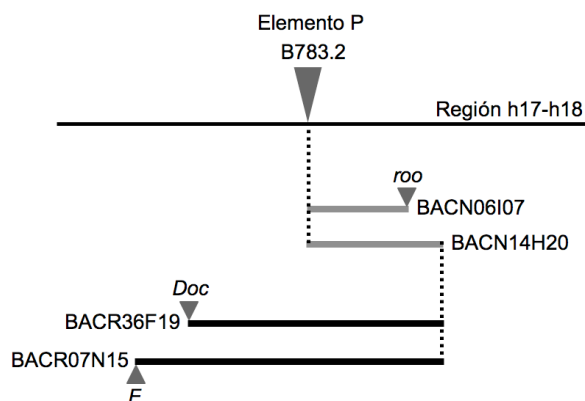


Figura 15. Análisis *in silico* empleado en la identificación de BACs de la región de inserción del elemento P B783.2.

Para comprobar que los BACs identificados pertenecían a la región h17-h18 se realizaron hibridaciones *in situ* sobre cromosomas prometafásicos de *D. melanogaster* empleando el DNA de los BACs como sonda. La figura 16 muestra que los BACs BACR07N15 y BACR36F19 localizan en la región h17. De este modo, el análisis realizado *in-silico* partiendo de los datos obtenidos en el estudio de la impronta genética del cromosoma Y permitió acceder a la región h17.

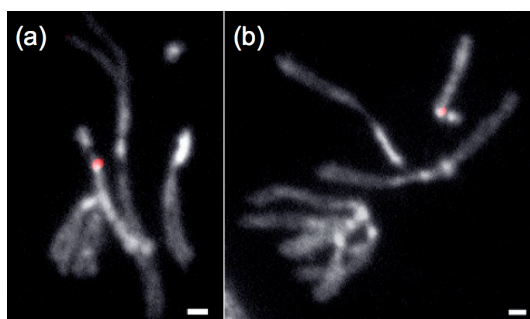


Figura 16. Hibridaciones *in situ* sobre cromosomas metafásicos.

Como sonda, se ha empleado (a) el BACR07N15 y (b) el BACR36F19. En ambos casos se observa una única señal en h17.

4.1.2.2. La región h17 contiene un gran palíndromo.

En colaboración con el WTSI se secuenció el BACR07N15, por ser el de mayor tamaño de los dos obtenidos. La secuencia del BAC, con número de acceso CU457433, ha sido analizada mediante BLAST y “Dotplot” y, posteriormente, anotada manualmente (ver Anexo II).

Brevemente, la secuencia obtenida corresponde a fragmentos procedentes de regiones eucromáticas, entremezcladas con muchos transposones degenerados. El análisis en más detalle muestra que el BACR07N15 está compuesto por varias duplicaciones segmentales que han sufrido amplificaciones posteriores, deleciones, inversiones e inserciones de elementos transponibles. Las duplicaciones segmentales proceden de la región 11B del cromosoma X y de las regiones 35B y 38C, del brazo largo del cromosoma 2. En particular, la duplicación “X11B” contiene el gen CG12717 y la mitad del gen *ade5*; la duplicación “2L35B” contiene el primer intrón del gen *Mst35B* (que codifica para una protamina) y la duplicación “2L38C” contiene secuencias del transposón INE. La figura 17 muestra la representación esquemática de una posible deconstrucción de la evolución experimentada por las secuencias de la región de h17 clonadas en el BAC. Así, tras la amplificación de la región que contenía las duplicaciones segmentales, se insertaron varios elementos transponibles: *diver2*, en primer lugar (pues se encuentra a su vez delecionado y amplificado) y *F*, *Idefix*, *aurora*, *Doc*, *roo* y *gypsy6* posteriormente. También se aprecia que, en distintos momentos, tuvieron lugar deleciones e inversiones. A continuación se formó un gran palíndromo, a partir de una región de unas 90 kb. Posterior a la formación de ese palíndromo, se insertaron otros transposones como *F*, *Idefix* y *opus*.

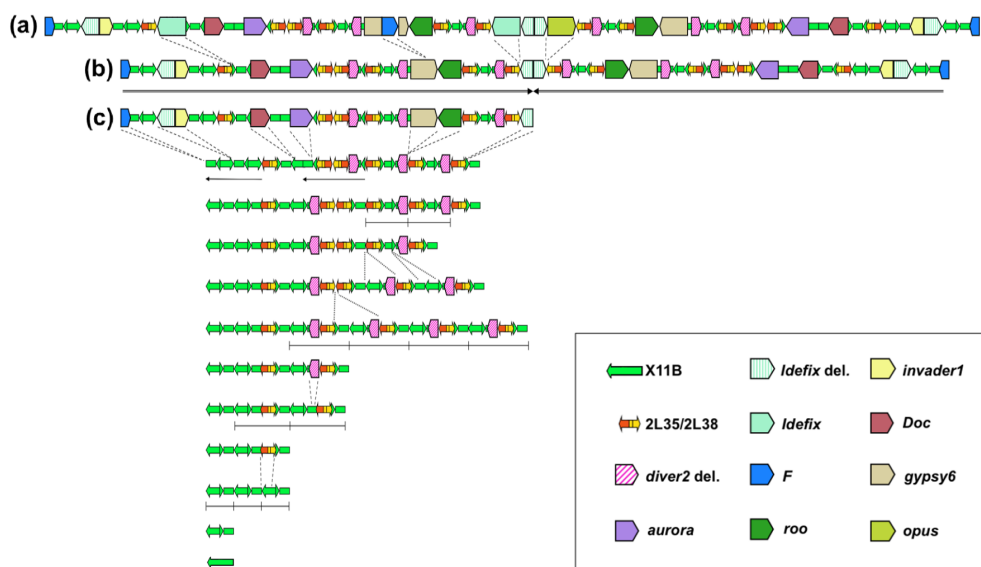


Figura 17. Deconstrucción de la secuencia palindrómica presente en el BACR07N15.

(a) Representación esquemática de los componentes del palíndromo. (b) Formación del palíndromo. Los brazos del palíndromo se indican con flechas horizontales convergentes. (c) Posibles pasos evolutivos que dieron lugar a la región que formó el palíndromo. Las amplificaciones e inversiones se indican mediante líneas horizontales en tándem y flechas horizontales, respectivamente. También se muestran las deleciones e inserciones. El número de acceso en el Genbank para la secuencia del BACR07N15 es CU457433.

La figura 18 muestra el “Dotplot” de la regiones palindrómicas, en el que se aprecia que la discontinuidad de la línea de puntos diagonal se debe a las inserciones posteriores a la formación del palíndromo de los elementos *ldefix* y *F*.

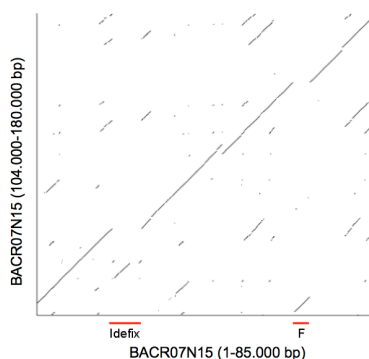


Figura 18. “Dotplot” de la primera mitad de la secuencia del BACR07N15 comparada contra la segunda mitad.

La falta de continuidad en la diagonal se deben a las inserciones de transposones.

4.1.2.3. El palíndromo de la región h17 abarca más de 200 kb

Hasta la fecha, nunca se había encontrado un palíndromo tan grande en el genoma de *D. melanogaster*. En el BACR07N15 además de la región palindrómica existen otras 18 kb con secuencias semejantes a las encontradas en el palíndromo. Este hecho sugería que el palíndromo continúa mas allá de lo detectado en este clon. Por este motivo, se decidió identificar otros clones que contuviesen las secuencias presentes en el palíndromo, para, de esta forma, obtener una estimación de su tamaño total.

Para proceder a la identificación de nuevos clones se diseñaron los cebadores 07N15-Fw y 07N15-Rv y se obtuvo, mediante PCR, una sonda que incluía el pseudo-intrón del gen *Mst35B*. Las condiciones de reacción se detallan a continuación (en la tabla 7).

Tabla 7. Cebadores y condiciones de la PCR para obtener la sonda del pseudo- intrón *Mst35B*.

<u>Cebadores</u>			
07N15-Fw: 5' TTTGATATCAACCCTTGTGG3'			
07N15-Rv: 5' CGTAGAGTGAAAGTATGTTG3'			
<u>Programa de PCR</u>			
Desnaturalización inicial	94°C, 5 min	}	30 ciclos
Desnaturalización	94°C, 1 min		
Anillamiento	52°C, 1 min		
Extensión	68°C, 1 min		
Extensión final	68°C, 10 min		

Tras analizar el producto de PCR obtenido en un gel de agarosa, se observaron dos bandas de tamaños de 1,1 y 1,3 kb. Ambas bandas se correspondían con el producto esperado, pues existe una delección de alrededor de 260 bases en algunas repeticiones del pseudo-intrón. Por lo tanto, tras la purificación del producto de PCR, éste fue marcado radioactivamente y

utilizado para hibridar las genotecas RPCI-98, CHORI221 y CHORI223. Los veinte clones que dieron señal de hibridación fuerte se muestran en la tabla 8, de la que se han excluido aquellos clones que daban una señal de hibridación débil, pues la secuencia de sus extremos mostró que derivaban de la región eucromática 35B donde se encuentra el gen *Mst35B*.

Tabla 8. Clones que hibridaron con la sonda del pseudo-intrón *Mst35B*.

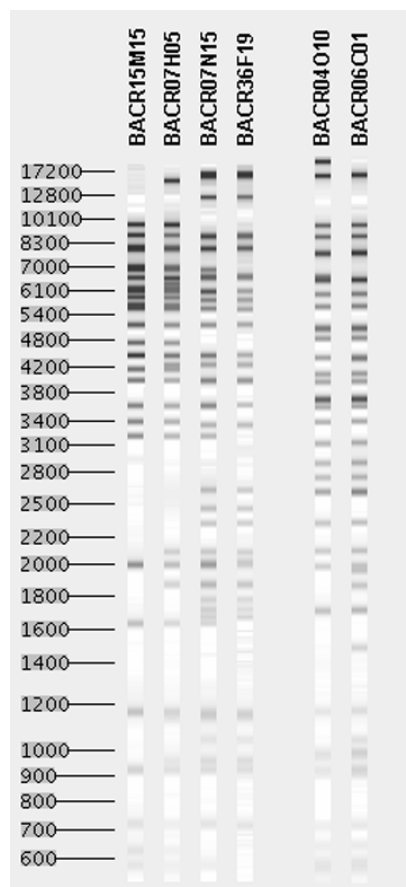
BACs	Ψ -Intrón <i>Mst35B</i>
BACR04O10	++++
BACR06C01	++
BACR07H05	+++
BACR07N15	+++++
BACR15M15	+++
BACR20C16	++++
BACR35N10	++
BACR36F19	++++
BACR41P19	++++
BACR42K20	+++
BACR45A24	+++
CH221-04I16	+++
CH221-29J24	+++
CH223-06C13	+++
CH223-07A15	++++
CH223-22J24	++++
CH223-28P23	+++
CH223-29N24	++
CH223-33K11	++
CH223-45L15	++

A partir del DNA de los clones obtenidos en el rastreo de las genotecas, se compararon sus digestiones con las enzimas *EcoRI* y *HindIII*, mediante la herramienta iCE (“Internet Contig Explorer”), del Centro de Ciencias del Genoma “Michael Smith” de Canadá. A pesar de que no todos los clones que dieron positivos con la sonda del pseudo-intrón *Mst35B* habían sido analizados en este Centro, se pudo comprobar que, al menos seis de los procedentes de la genoteca RPCI-98 si que lo estaban. Además, se observó que algunos de estos BACs solapaban entre si. El programa iCE tenía distribuidos estos seis clones en tres “contigs” independientes: el 586, 634 y 635. Los clones que pertenecen a cada “contig”, así como el tamaño medio estimado del “contig” se detallan en la tabla 9.

Tabla 9. “Contigs” que contienen clones positivos con la sonda del pseudo-intrón *Mst35B*.

Nº Contig	Clones	Tamaño estimado (kb)
586	BACR07N15 BACR36F19	170
634	BACR04O10 BACR06C01	169
635	BACR07H05 BACR15M15	149

La figura 19 muestra las imágenes de las digestiones de estos seis BACs, donde se observan las bandas coincidentes entre algunos de ellos. Es importante destacar que en las imágenes de las digestiones no se percibe la presencia de satélites en ninguno de los clones (que se observarían como bandas de mayor intensidad que la esperada).

**Figura 19. Análisis de restricción de 6 BACs digeridos con la enzima *EcoRI*.**

Haciendo uso de toda la información recabada sobre la región h17 se ha construido un “contig” de unas 500 kb, en las cuales no se halló ningún satélite que pudiese estar participando de la función centromérica (ver figura 20).

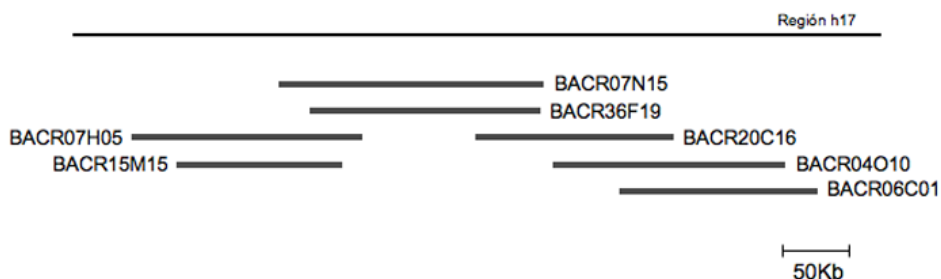


Figura 20. “Contig” de la región h17 del cromosoma Y.

En la construcción del “contig” se contrastó la información obtenida mediante enzimas de restricción con los datos derivados de comparar las secuencias existentes de los extremos de los BACs contra la secuencia del BACR07N15. El alto grado de amplificación que sufrió esta región, y la presencia de un palíndromo, dificultaron enormemente esta tarea. Ambos extremos del BACR36F19 dieron homología con secuencias del BACR07N15 y el BACR36F19 era de menor tamaño, por lo que se concluyó que estaba incluido en el BACR07N15, como se observa fácilmente en el análisis de restricción. Además uno de los extremos del BACR07H05, del BACR15M15 y del BACR20C16, daban alta homología con distintas regiones del BACR07N15, por lo que se pudo obtener la posición exacta de los BACs respecto de éste. Por otra parte, el BACR07H05 y el BACR15M15 solapaban entre sí, siendo el primero mayor que el segundo, y, por lo tanto, la totalidad de la secuencia presente en el BACR15M15 está incluida en la del BACR07H05. Mas aún, tras el análisis de restricción, se sabe que el BACR04O10 y el BACR06C01 también solapaban entre sí, sin que ninguno incluyera completamente la secuencia del otro, y que además, si bien tienen secuencias parecidas a las del BACR07N15 (como se observa al hacer un BLAST con la secuencia del extremo T7 del BACR04O10), no tienen solapamiento alguno con el mismo. Lamentablemente, al analizar las secuencias de los extremos de los BACs que habían dado

positivo en la hibridación con la sonda Mst35B-Ψ, no se obtuvo ninguna secuencia única, o de otro BAC, que permitiese seguir avanzando en el conocimiento de la región h17.

4.2. El centrómero del cromosoma 3 de *D. melanogaster*

4.2.1. Análisis de la región centromérica del cromosoma 3

4.2.1.1. La proteína centromérica CID colocaliza con el satélite dodeca en fibras de cromatina

En la actualidad hay dos formas de corroborar que una secuencia determinada está formando cromatina centromérica, y por lo tanto, es parte de un centrómero funcional. Una de las opciones es mediante inmunoprecipitación de la cromatina (técnica conocida como ChIP), empleando anticuerpos contra CENP-A y la otra, mediante colocalización de CENP-A con la secuencia en cuestión en fibras de cromatina extendida. El primer método fue abordado en *D. melanogaster* por el grupo del Prof. Karpen. Se realizó un ensayo de inmunoprecipitación de la cromatina con el anticuerpo de CID (homólogo en *Drosophila* de la CENP-A humana), pero las secuencias de DNA que se obtuvieron fueron analizadas sin que se llegase a nada concluyente. Es más, recientemente se ha comprobado que algunas de las secuencias que se obtuvieron del cromosoma 3 corresponden a regiones eucromáticas. Los datos derivados de este análisis, realizado en 2001, nunca fueron publicados, por lo que cabe concluir que la técnica falló, debido a que la naturaleza repetida de los satélites dificulta la delimitación precisa del DNA centromérico. En vista de estos resultados, en este trabajo se decidió abordar el estudio mediante inmunofluorescencia e hibridación *in situ* con sondas fluorescentes. Los problemas de este enfoque radican en que hibridando cromosomas prometafásicos no se obtiene suficiente resolución y en que la heterocromatina pericentromérica no politeniza, quedando los centrómeros agregados en el cromocentro del cromosoma politénico (Zhang y Spradling, 1995). Sin embargo, el grupo del Prof. Zhimulev ha desarrollado recientemente mutantes dobles para los genes *Su(var)3-9* y *SuUR*, que codifican para una metiltransferasa de histonas y para un regulador negativo de la replicación del DNA, respectivamente (Andreyeva et al., 2007). Debido a la fuerte supresión

de la baja replicación habitual de la heterocromatina en los cromosomas politénicos, los dobles mutantes presentan el bandeo característico de estos cromosomas también en la región centromérica del cromosoma 3.

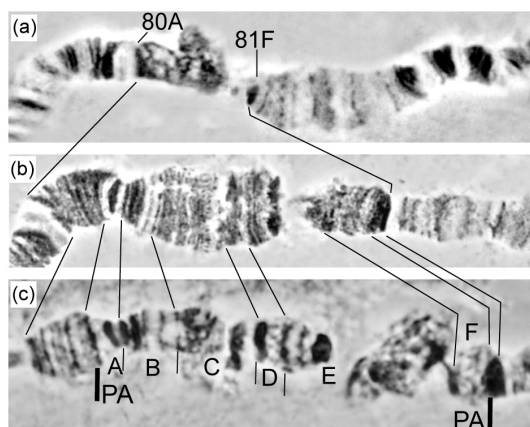


Figura 21. Región centromérica del cromosoma politénico 3 de *D. melanogaster* (a) en la cepa silvestre, (b) en el mutante *SuUR*, y (c) en el doble mutante *Su(var)3-9* y *SuUR*. PA se refiere a Plato Atlantis, y éste ha sido subdividido en las regiones PAA-PAF.

A esta nueva región politenizada se la ha designado como Plato Atlantis (PA), y su bandeo ha sido caracterizado y nombrado con letras de la A a la F (PAA-PAF, ver figura 21) (Andreyeva et al., 2007). Gracias a la colaboración establecida con este grupo, se pudo observar, mediante inmunofluorescencia e hibridación *in situ*, que tanto la proteína centromérica CID como el satélite dodeca localizaban en PAE (ver figura 22).

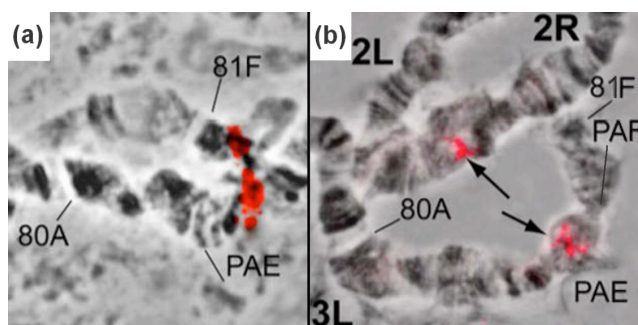


Figura 22. Cromosomas politénicos del doble mutante *Su(var)3-9* y *SuUR*. (a) Hibridación con la sonda del satélite dodeca. (b) Inmunofluorescencia con el anticuerpo anti-CID. En ambos casos, la señal aparece en la región PAE.

Una vez observada la localización en cromosomas politénicos, se decidió aumentar el grado de resolución mediante inmunofluorescencia e hibridación sobre fibras de cromatina extendida. La figura 23 muestra la colocación de CID (inmunofluorescencia, en verde) y la sonda dodeca (FISH, en rojo) en una fibra de cromatina, obtenida a partir de células S2 de *D. melanogaster*. Esta colocación muestra que el satélite dodeca forma cromatina centromérica.

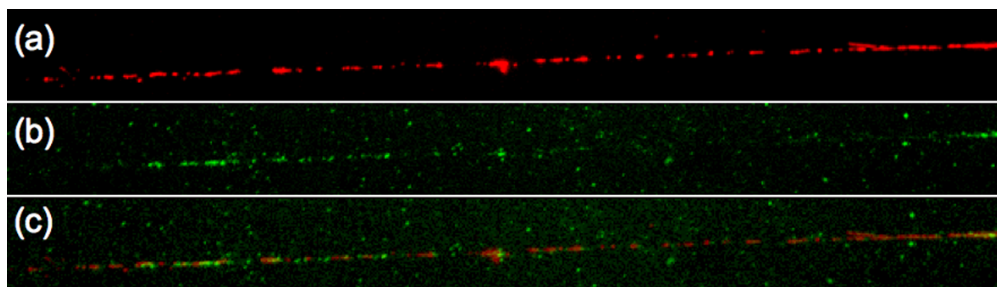


Figura 23. Hibridación *in situ* e inmunofluorescencia sobre fibras de cromatina.

(a) Sonda del satélite dodeca (rojo). (b) Anticuerpo contra CID (verde). (c) Montaje de las señales de FISH e inmunofluorescencia.

4.2.1.2. Secuencias relacionadas con el espaciador intergénico ribosomal en h53

Tras la colocación de CID con el satélite dodeca se decidió realizar un estudio estructural de la región h53 en profundidad, con la intención de caracterizar todas las secuencias existentes en esta región. Se comenzó rastreando tres genotecas de BACs (RPCI-98, CHORI221 y CHORI223) con una sonda del satélite dodeca. En análisis previos de la heterocromatina de *D. melanogaster* realizados en el laboratorio, se habían rastreado estas mismas genotecas con sondas de *HeT-A* y *Circe*. *HeT-A*, como ya se ha mencionado, es un retrotransposón telomérico. *Circe*, por su parte, es un transposón LTR que aparece asociado mayoritariamente a la heterocromatina constitutiva (Losada et al., 1999a). Además, a partir de análisis previos, se había obtenido un mapa físico de la región centromérica del cromosoma 3 (Losada et al., 2000), que se ha empleado en este trabajo como punto de partida. La figura 24 muestra la localización de *Circe*, *HeT A* y el satélite dodeca en la región centromérica del cromosoma 3 de *D. melanogaster*.

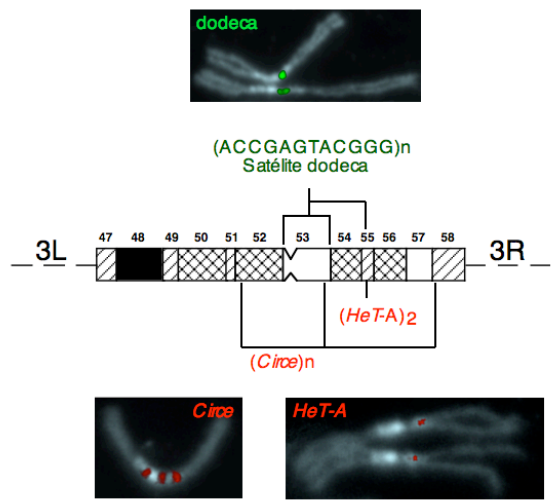


Figura 24. Localización de *Circe*, *HeT-A* y del satélite dodeca en la heterocromatina del cromosoma 3 de *D. melanogaster*.
 Se muestran las hibridaciones *in situ* con cada una de las sondas.

A continuación se muestran las tablas 10.a-c en las que se han recopilado los datos de estas tres hibridaciones, y se han ordenado de modo que la tabla 10.a detalla los clones que dieron positivos tanto para la sonda de dodeca como para la de *Circe*, la tabla 10.b los que dieron positivo para dodeca y *HeT-A* y, por último, la tabla 10.c los que únicamente dieron positivos con la sonda de dodeca, pertenecientes al cromosoma 3.

Tabla 10. Clones que hibridan con las sondas (a) de dodeca y *Circe*, (b) de dodeca y *HeT-A* y (c) solamente con la sonda de dodeca.

(a)

BACs	dodeca	Circe
BACR03N05	+++	+++
BACR09A22	++++	+++
BACR12I02	++++	++
BACR13P24	+++++	+++
BACR16A01	+++	++
BACR19P07	+++++	++
BACR20F21	++++	++
BACR22C05	+++++	+
BACR23C02	++++	++
CH221-25D13	++++	+
CH221-25K16	+++++	++
CH223-06A08	++++	+++
CH223-06E11	++	++
CH223-12J10	++++	++
CH223-13D04	++	++
CH223-19A02	++++	++
CH223-21H21	+++	++
CH223-21O18	+++	++
CH223-25I15	+++	++
CH223-29D12	+++	++
CH223-46O09	++++	+++
CH223-48J18	++++	++

(b)

BACs	dodeca	HeT-A
BACR04N14	++++	+++
BACR04P01	+++	++++
BACR10K06	+++	+
BACR17D04	++++	+++
BACR17E23	++++	++++
BACR19G05	++++	++
BACR23A16	++++	++++
BACR23G18	+++	+++
BACR24E04	++++	+++
BACR24G14	++++	++
BACR24P14	+	++
BACR48H02	++++	+++
BACR48I03	++	+++
CH221-07O07	++++	++
CH221-20E13	++++	++
CH221-22O22	++++	+++
CH221-33K19	+++	+++
CH223-32N13	++++	+++
CH223-42E08	+++	+++

(c)

BACs	dodeca	BACs	dodeca
BACR04C18	++	CH223-07F15	++++
BACR06I12	++	CH223-10A17	++++
BACR06P11	++	CH223-11E06	+++
BACR07C06	++	CH223-11M22	+++
BACR08A23	++	CH223-13P21	++
BACR13C23	+++	CH223-14L06	++
BACR19E12	++	CH223-16M07	+
BACR19O11	++	CH223-19B15	+++
BACR19O13	++	CH223-20N14	+++
BACR21F02	+++	CH223-25K12	+++
BACR21J06	++	CH223-26G16	++
BACR25P13	+++	CH223-27D01	+++++
BACR38D06	++	CH223-28G12	++
BACR38N04	++	CH223-29K11	+
CH221-01A17	+	CH223-31J03	+++
CH221-20O01	++++	CH223-34J07	++++
CH221-27P10	+++	CH223-37A15	++++
CH221-29J09	++	CH223-37D12	+++
CH223-02L21	+++	CH223-38K16	+++
CH223-03G09	++++	CH223-39I07	+++
CH223-03N05	+++	CH223-39J07	++
CH223-04D21	++++	CH223-41G08	+++
CH223-04H09	++++	CH223-41M22	+++
CH223-06P02	+	CH223-41N24	++++
CH223-06P22	++		

Tras el rastreo de las genotecas con el satélite dodeca, se eligieron 80 clones que daban señal de hibridación fuerte. Los extremos del inserto de cada uno de estos BACs fueron secuenciados, y sus lecturas fueron analizadas contra las bases de datos de *D. melanogaster*. Entre las lecturas se obtuvieron homologías con secuencias de transposones, duplicaciones segmentales (2L 29C: *Akap200*), secuencias simples repetidas (CATATATA)_n y secuencias relacionadas con el espaciador intergénico ribosomal (IGSr), entre otras. Sorprendentemente, entre los 80 clones secuenciados sólo hubo dos casos en los que las lecturas de ambos extremos dieron homología con dodeca. Por otra parte, el hecho de que apareciesen secuencias relacionadas con el espaciador intergénico ribosomal en la región centromérica del cromosoma 3 era completamente inesperado, puesto que éstas sólo se conocían en las regiones del organizador nucleolar (NOR) de los cromosomas X e Y. El espaciador intergénico ribosomal (IGS) está formado por repeticiones de 240 bp que, según se ha descrito, están implicadas en el apareamiento de los cromosomas X-Y en meiosis (Ren et al., 1997). La búsqueda en bases de datos de secuencias que contuviesen IGSr proporcionó el ensamblado CP000331, del Release 5.1 de la heterocromatina de *Drosophila* (Smith et al., 2007). Este ensamblado contenía el IGSr unido directamente a dodeca, pero se desconocía su localización en el genoma. En las búsquedas en bases de datos se obtuvo también homología con secuencias en las que IGSr aparecía junto al satélite (AATAACATAG)_n (satélite 10bp, en adelante). El satélite 10bp fue caracterizado por (Lohe et al., 1993), quienes lo localizaron en h37, próximo al centrómero del cromosoma 2, y en la región h48 del cromosoma 3.

En vista de este resultado se planteó si sería posible encontrar IGSr asociado también al centrómero del cromosoma 2. En primer lugar se realizaron hibridaciones *in situ* sobre cromosomas prometafásicos empleando como sonda un BAC que contenía dodeca e IGSr. La figura 25.a muestra que el BACR31J03 sólo hibrida en la región centromérica del cromosoma 3 de *D. melanogaster*. Sin embargo, como se desconocía la cantidad de IGSr presente en el BAC, se decidió obtener una sonda de IGSr “pura” de esta región. A partir de un fragmento de 6 kb *EcoRI-EcoRI* del BACR31J03 se amplificó mediante

PCR la sonda IGSr (5,6 kb). La secuencia de los cebadores IGSr-Fw e IGSr-Rv y las condiciones de la reacción de PCR están detallados en tabla 11. La figura 25.b muestra que al usar la sonda IGSr (5,6 kb) tampoco apareció señal alguna en la región centromérica del cromosoma 2. Como podía ser que el IGSr del cromosoma 2, en el caso de existir, tuviese una secuencia algo divergente, se decidió hibridar en condiciones menos restrictivas, bajando la temperatura de hibridación, así como utilizar la técnica PRINS (del inglés, *primed in situ labeling*). En estos experimentos, se pudo ver hibridación cruzada con el NOR del cromosoma X, pero siguió sin verse señal en el cromosoma 2.

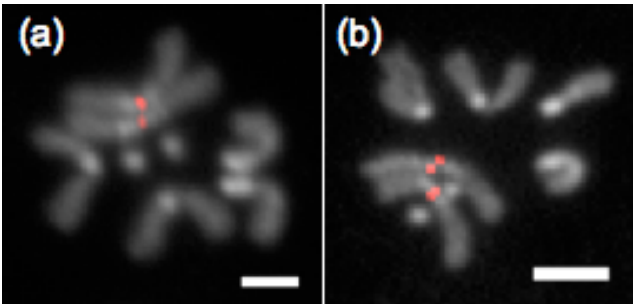


Figura 25. Hibridaciones *in situ* sobre cromosomas metafásicos.
Se ha empleado como sonda (a) el BACR31J03 y (b) el fragmento de 5,6 kb de IGSr. La barra representa 2 micras.

Tabla 11. Cebadores y condiciones de la PCR para obtener la sonda IGSr.

Cebadores

IGSr-Fw: 5' TGGCAGCGTTTTAAGGGATG 3'

IGSr-Rv: 5' TAAGACGCCTGCAGAGAACG 3'

Programa de PCR

Desnaturalización inicial	94°C, 1 min	} 35 ciclos
Desnaturalización	96°C, 20 seg	
Anillamiento	55°C, 1 min	
Extensión	68°C, 7 min	
Extensión final	68°C, 10 min	

Dado que de ninguna de las formas abordadas en el intento de encontrar secuencias relacionadas al IGS en el centrómero del cromosoma 2 dieron un

resultado positivo, pero existían secuencias en las que IGSr se encontraba a continuación del satélite 10bp, se decidió realizar hibridaciones *in situ* dobles, empleando simultáneamente la sonda de IGSr, previamente descrita, y una sonda del satélite 10bp. La secuencia del oligonucleótido marcado que se empleó como sonda se muestra en la tabla 12. Debido al elevado contenido en AT de la sonda del satélite 10bp, la hibridación se realizó en condiciones poco restrictivas (25°C). La figura 26 muestra cromosomas prometafásicos con un bajo grado de condensación, donde es fácil observar que la sonda del satélite 10bp hibrida en h52, próxima a la señal de IGSr. Sorprendentemente no se observó que la sonda del satélite 10bp marcara la región h48 del cromosoma 3, donde previamente había sido localizado este satélite (Lohe et al., 1993). En esta figura se puede ver que, debido a la baja temperatura de hibridación requerida para la sonda del satélite 10bp, la sonda de IGS también hibrida con el NOR del cromosoma X.

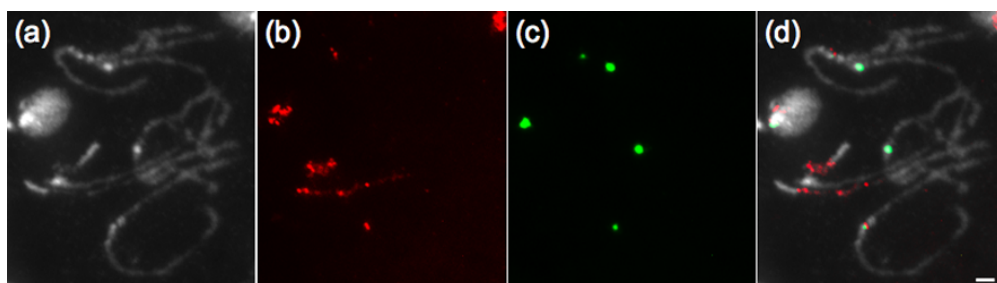


Figura 26. Hibridación *in situ* de (a) cromosomas prometafásicos.

Hibridación *in situ* de (a) cromosomas prometafásicos, con las sondas (b) de IGSr (rojo) y (c) del satélite 10bp (verde). (d) Montaje con ambas sondas. La barra representa 2 micras.

Para resolver si el IGSr unido a dodeca y el IGSr unido a 10bp provenían de dos zonas distintas de la región centromérica del cromosoma 3, se recurrió a realizar hibridaciones *in situ* sobre cromosomas politénicos de los mutantes dobles para los genes *Su(var)3-9* y *SuUR*, previamente descritos, con sondas de IGSr y satélite 10bp. Gracias a la resolución que proporciona el doble mutante en la región conocida como Plato Atlantis, se pudo comprobar que las secuencias de IGSr están presentes en dos localizaciones distintas de la región centromérica del cromosoma 3: una de ellas entre el satélite 10bp y el principio

del primer bloque de dodeca (PAE 3-4), y la otra, posiblemente, entre los dos bloques de dodeca (PAE 5-6), como se observa en la figura 27.

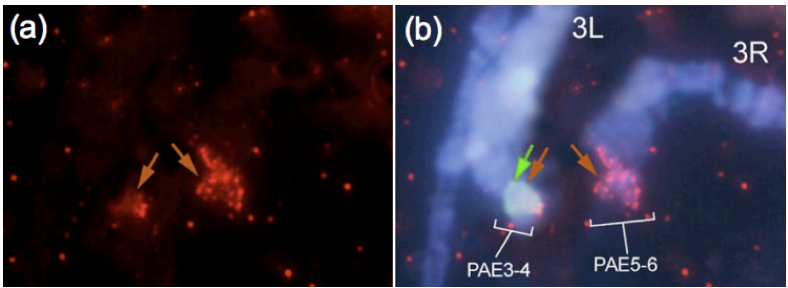


Figura 27. Hibridación *in situ* sobre cromosomas politénicos del doble mutante *Su(var)3-9* y *SuUR*.
(a) Señal de la sonda IGSr, (b) Montaje de las señales de las sondas de IGSr (rojo) y del satélite 10bp (verde).

Tabla 12. Secuencia de oligonucleótidos marcados fluorescentemente.

Sonda	Secuencia	Marcaje
Satélite 10bp	5'AATAACATAGAATAACATAGAATAACATAGAA TAACATAGAATAACATAG 3'	3' fluoresceína 5' cy3
Satélite 15bp	5'AACATGTTCTGTTCTGAACATGTTCTGTTCTGAA CATGTTCTG 3'	5' cy3
Satélite dodeca	5'CCCGTACTGGTCCCGTACTGGTCCCGTACTC GGTCCCGTACTCGGT3'	3' fluoresceína 5' cy3

4.2.1.3. El satélite de 10bp está presente en la región proximal de h52

El inesperado hallazgo del satélite 10bp en h52, y no en h48, implicaba que éste debería encontrarse próximo al satélite dodeca, como se observó mediante hibridaciones *in situ* dobles con sondas fluorescentes de los satélites dodeca y 10bp (ver figura 28).

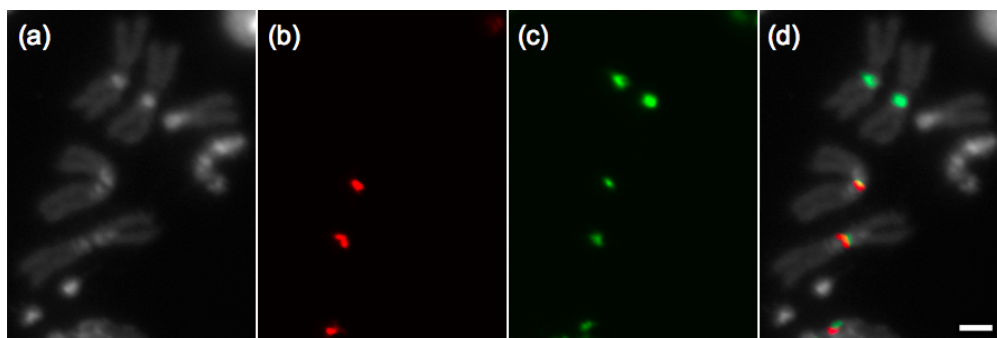


Figura 28. Hibridación *in situ* de cromosomas metafásicos.

Hibridación *in situ* de (a) cromosomas metafásicos, con las sondas de los satélites (b) dodeca (rojo) y (c) 10bp (verde). (d) Montaje de los cromosomas con ambas sondas. La barra representa 2 micras.

Losada y colaboradores realizaron un mapa físico de 1,5 Mb de la región centromérica del cromosoma 3 de *D. melanogaster*, utilizando DNA de la cepa isogénica para el cromosoma 3 *red e* (Losada et al., 2000). Para el análisis de restricción del DNA de alto peso molecular utilizaron enzimas que cortan el DNA con baja frecuencia (como *BamHI*, *BssHI* y *SwaI*). En ese estudio propusieron la existencia de un satélite de DNA, de naturaleza desconocida, flanqueando el satélite dodeca. Para comprobar que el satélite 10bp correspondía con el predicho por Losada y colaboradores se hibridaron filtros que contenían DNA de alto peso molecular digerido con las enzimas *BamHI*, *BssHI* y *SwaI* con las sondas del satélite dodeca y del satélite 10bp. La figura 29 muestra que la banda de 1,2 Mb en el DNA digerido con *BssHI* y la de ~430 kb en la digestión doble con *SwaI* y *BamHI* hibridan tanto con la sonda de dodeca como con la del satélite 10bp. Como se conoce la localización de estos fragmentos de DNA en el mapa físico de la región se pudo determinar que el satélite 10bp flanquea al bloque de dodeca próximo a h52.

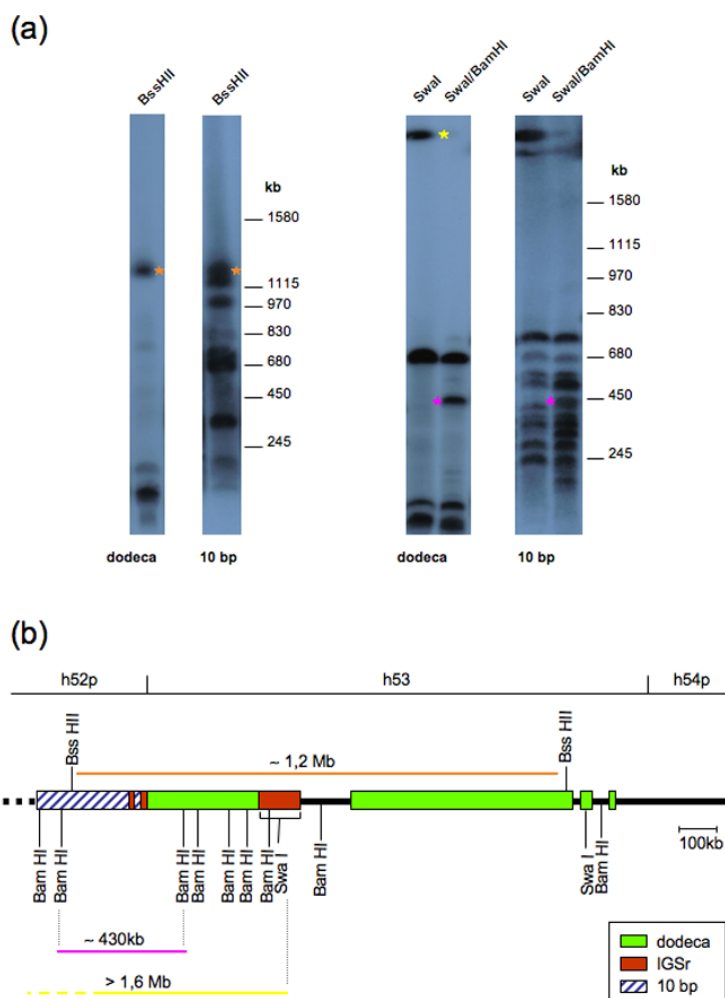


Figura 29. Hibridaciones en filtro con sondas de los satélites dodeca y 10bp.

(a) Hibridaciones en filtro con sondas de los satélites dodeca y 10bp del DNA de alto peso molecular digerido con las enzimas de restricción *BssHII*, *SwaI* y *BamHI*. (b) Representación esquemática del mapa de restricción de la región h52p-h54p. El fragmento *BssHII*-*BssHII*, de ~1,2 Mb aparece indicado en naranja; el fragmento *SwaI*-*SwaI*, de > 1,6 Mb, en amarillo, y el fragmento *SwaI*-*BamHI*, de ~430 kb, en fucsia.

Por último, el rastreo de las tres genotecas de BACs con la sonda del satélite 10bp permitió identificar 32 clones (ver tabla 13). Entre ellos, se encontró que los clones CHORI223-10H11 y CHORI223-19F18 se encuentran en la parte proximal del ensamblado CP000217 de la región h52. Como ambos clones dieron una señal de hibridación muy fuerte se estima que estos clones contienen parte del bloque de satélite 10bp presente en h52p.

Tabla 13. Clones que hibridan con la sonda del satélite 10bp.

BACs	10bp	BACs	10bp
BACR06H21	+	CH223-10H11	+++
BACR06L04	++	CH223-13H01	+++
BACR12B13	+	CH223-15L17	+
BACR19I21	+	CH223-18O23	+
BACR26L11	++	CH223-19F18	++++
BACR28O08	++++	CH223-25E18	+++
BACR30G20	+	CH223-25L21	+
BACR31O01	+	CH223-27H20	+++
BACR34A01	+	CH223-28P19	+++
BACR35L16	+	CH223-30D16	+
BACR42H13	+	CH223-30K04	+
BACR44E02	+	CH223-33D04	+
BACR44G02	+	CH223-33E07	+++
CH221-16H11	+	CH223-33L06	++
CH221-25A23	+	CH223-41C03	++
CH221-46C07	+	CH223-46K17	++

4.2.1.4. ¿Existe el satélite de 15bp en *D. melanogaster*?

En las bases de datos de *D. melanogaster* se encontraron secuencias que contenían el satélite de 15bp, $-(AACATGTTCTGTTCG)_n-$. Este satélite, que es un componente mayoritario de la heterocromatina de los cromosomas 2 y 3 de *D. simulans*, *D. mauritiana* y *D. sechellia* (Lohe y Roberts, 1988), no había sido descrito previamente en *D. melanogaster*. Con la intención de localizarlo, se realizaron hibridaciones *in situ* sobre cromosomas metafásicos de *D. melanogaster*. En paralelo, se hibridaron cromosomas metafásicos de *D. simulans* y *D. mauritiana*, con sondas de los satélites 15bp, 10bp y dodeca. Las secuencias de los oligonucleótidos marcados empleados como sondas para estas hibridaciones se muestran en la tabla 12. La figura 30 muestra las hibridaciones dobles obtenidas para cada especie. Tanto *D. simulans* y *D. mauritiana* presentan señales de hibridación en la región centromérica de los cromosomas 2 y 3 con las sondas de dodeca y del satélite 15bp, pero no con el satélite de 10bp. En el caso de *D. melanogaster*, las sondas de los satélites 10bp y dodeca dan las señales esperadas, mientras que, la sonda del satélite

15bp no da señal alguna. Este dato negativo indica que la sensibilidad de la técnica no es la adecuada para detectar un bajo número de repeticiones de este satélite.

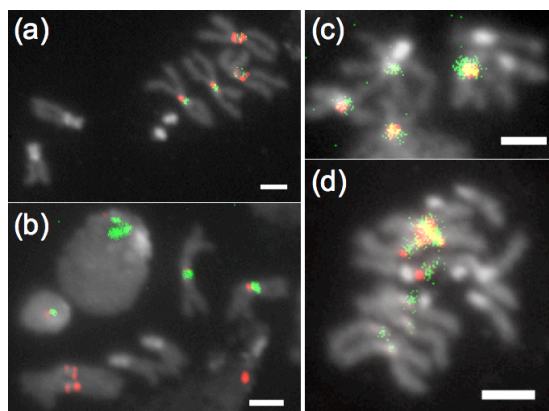


Figura 30. Hibridaciones *in situ* sobre cromosomas metafásicos de *D. simulans* y *D. mauritiana*.

Hibridaciones *in situ* con las sondas de los satélites dodeca (verde) y 15bp (rojo) sobre cromosomas metafásicos de (a, b) *D. simulans* y (c, d) *D. mauritiana*. La barra representa 2 micras.

4.2.1.5. Análisis de la secuencia obtenida de la región centromérica h53

A partir de la información recabada del mapa físico de h53, del rastreo de las genotecas con varias sondas, de la secuenciación de los extremos de 80 BACs que contenían dodeca y del análisis de restricción de ciertos BACs, se distribuyeron los 80 BACs entre las regiones h53-h56 (ver figura 31). Para ello, se consideró, en primer lugar, que los clones que también eran positivos con la sonda de *HeT-A* pertenecen a la región h55-h56, donde previamente se había descrito la presencia de este retrotransposón (Losada et al., 1999b). Los clones débiles con la sonda de dodeca y cuyos extremos mostraban homología con secuencias del ensamblado CP000221 correspondían a la región h54 (Hoskins et al., 2007). Losada y colaboradores detectaron la presencia del retrotransposón *Circe*, entremezclado con dodeca, en la zona proximal de h54 (Losada et al., 1999a). Por esto, los clones que dieron positivo con dodeca y *Circe* fueron asignados a la región h53-h54p. Los extremos de los clones que contenían secuencias de IGSr daban homología con el ensamblado CP000331, que a su vez tenía secuencias de dodeca, en las que aparecen dianas de

restricción para la enzima *Bam*HI. En base al mapa físico previo de la región, esto sólo sucede en el bloque de dodeca próximo a h52, por lo que estos clones fueron asignados a la isla entre los dos bloques principales de dodeca (h53L). El resto de los clones se asignaron a la región h53R. Entre estos, los clones que contenían secuencias de *Akap200* por un lado y los que contenían (CATATATA)_n por otro lado, se han localizado, basándose en el patrón de restricción con las enzimas *Swal* y *Hpa*I, en los extremos del bloque grande de dodeca, como muestra la tabla.

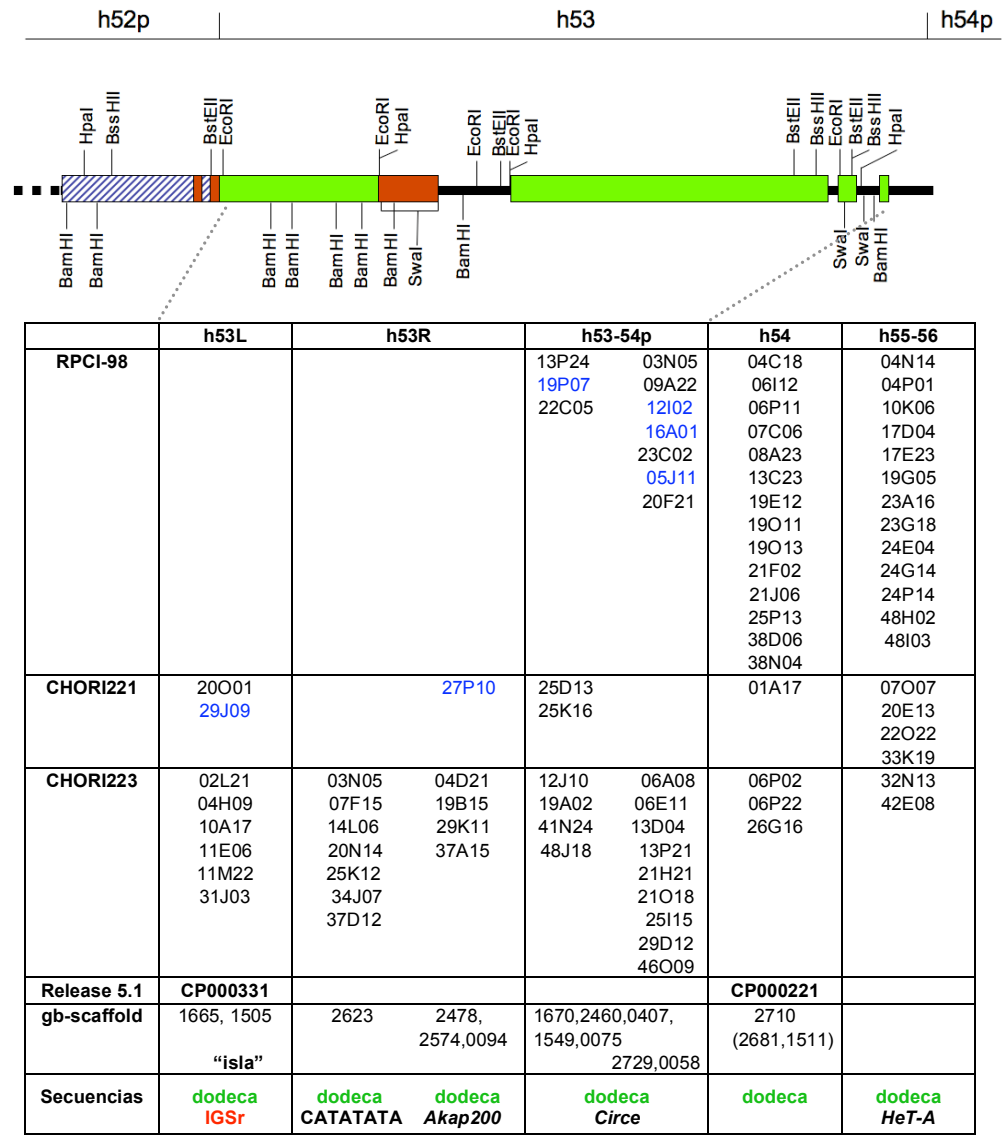


Figura 31. Clones procedentes de las genotecas de BACs que contienen el satélite dodeca, ordenados en la regiones h52p- h56. En la parte superior aparece un esquema del mapa físico de esta región centromérica.

Una vez asignados todos los clones y con la intención de obtener la máxima representación de la región h53 se secuenciaron completamente 4 BACs de la región h53-54p (BACR19P07, BACR16A01, BACR12I02 y BACR05J11), uno de h53R (CHORI221-27P10) y uno de h53L (CHORI221-29J09).

La secuenciación de los cuatro clones de la región h53-54p ha permitido obtener unas 400 kb de secuencia continua. Estos clones son: BACR19P07, BACR16A01, BACR12I02 y BACR05J11. Desafortunadamente, la secuencia del inserto del BAC CHORI221-27P10 no solapaba con el “contig” de 400 kb. El análisis de las secuencias mostró tramos de dodeca, interrumpidos por gran número de transposones degenerados (figura 32). Además se observa más secuencias de dodeca a medida que se avanza hacia h53. No se observa que entre los transposones haya una prevalencia en el tipo, pues siendo los más frecuentes los retrotransposones LTR, también se han encontrado elementos no-LTR, así como elementos TIR. Por otra parte, se han detectado duplicaciones segmentales derivadas de la región 2L-29C (*Akap200*) y del NOR del cromosoma X o Y (IGS).

Agavillando la información obtenida durante el análisis de la región centromérica del cromosoma 3, se puede alcanzar una visión bastante precisa de la organización de las secuencias presentes en esta zona. Al avanzar del brazo 3L hacia el centrómero, aparece, en primer lugar, un bloque del satélite 10bp, a continuación aparecen secuencias del IGSr entremezcladas con tramos del satélite 10bp. Después, aparece el bloque más pequeño de los dos bloques principales de dodeca. En la isla entre los dos bloques de dodeca hay secuencias IGSr seguidas de una región enriquecida en elementos transponibles. Por último, aparece el gran bloque de dodeca, que en su extremo próximo al brazo 3R presenta múltiples inserciones de transposones, así como la presencia de duplicaciones segmentales.

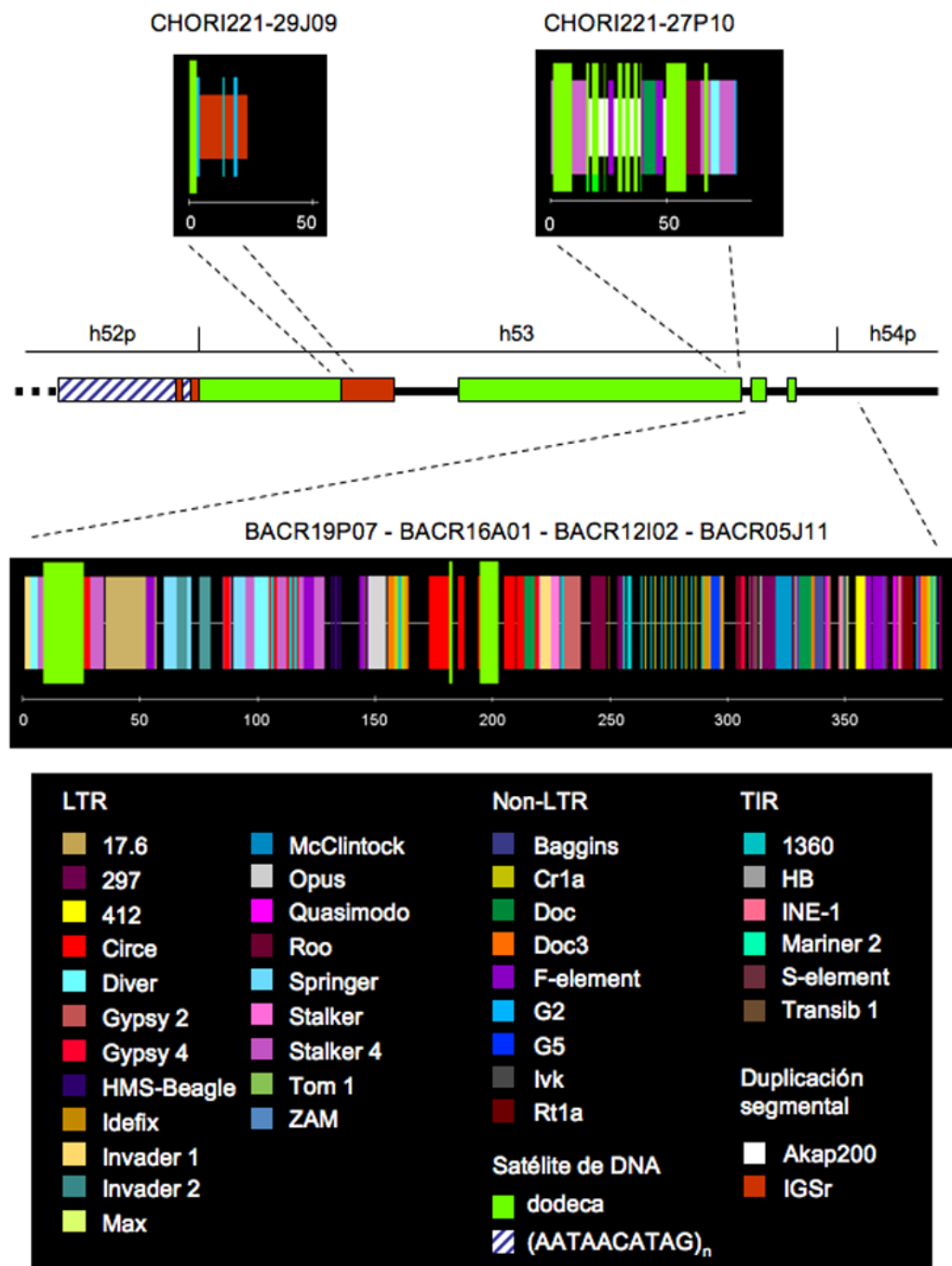


Figura 32. Representación de la anotación de la secuencia obtenida a partir de seis BACs de la región centromérica del cromosoma 3.

Los transposones han sido ordenados en la leyenda en función de su clasificación como LTR, no-LTR o TIR (Kaminker et al., 2002). Los números de acceso en el Genbank para las secuencias de los BACs BACR19P07, BACR16A01, BACR12I02, BACR05J11, CHORI221-27P10 y CHORI221-29J09 son, respectivamente, CU311183, CR942806, CR942807, CR942808, CU313318 y CU463787.

Discusión

5. DISCUSIÓN

5.1. La técnica de secuenciación

La posibilidad de secuenciar regiones de DNA altamente repetido mediante la estrategia basada en la secuenciación por transposición, presentada en esta tesis doctoral, representa un avance importante en el desarrollo de una técnica universal que permita la caracterización de los centrómeros de los cromosomas eucarióticos, así como de otras regiones heterocromáticas.

A pesar de la similitud observada en las secuencias obtenidas por el método nuevo aquí presentado y por las técnicas de “finalización de la secuencia” habitualmente empleadas por los grupos especializados del WTSI, es importante destacar todas las ventajas de la secuenciación por transposición que emplea la información de la posición del transposón. En primer lugar, si se compara el número de lecturas que se emplearon en cada caso para la consecución de la secuencia final, 933 lecturas en el CBMSO y 12993 en el WTSI, se observa que el WTSI necesitó un orden de magnitud mayor de lecturas. Esto, en términos económicos, resulta en un abaratamiento de la secuencia. Por otra parte, la posibilidad de emplear la técnica desarrollada sólo depende de la presencia de una diana *PI-SceI* o *I-SceI* en el vector de la genoteca de BACs en cuestión. Esto hace de ésta una técnica universal, pues no es preciso conocer la naturaleza de la secuencia problema, y en función de la misma decidir un protocolo de actuación. De este modo se ahorra en tiempo y esfuerzo, puesto que el abordaje será siempre el mismo. Todos estos motivos hacen de esta técnica una candidata ideal para automatizarla.

Es cierto que para que su automatización se lleve a cabo sería preciso modificar alguno de los pasos que en esta tesis se describen. Éste sería el caso de las técnicas de mapeo empleadas. En el desarrollo de la técnica se ha utilizado la electroforesis de campo pulsado para realizar la medición de los fragmentos generados tras la digestión del DNA de los clones con *PI-SceI*. Además, tras la electroforesis, el DNA del gel fue transferido a una membrana y posteriormente hibridado con sondas radioactivas. Este proceso es bastante laborioso y difícil de automatizar. Sin embargo, se ha visualizado una manera

de realizar el mapeo de la inserción del transposón apta para la automatización. La estrategia consistiría en realizar la medición del tamaño de los dos fragmentos mediante mapeo óptico (“optical mapping”, (Giacalone et al., 2000). Esta técnica serviría también para distinguir entre los dos fragmentos. Así, en lugar de emplear sondas radioactivas que hibriden en uno de los fragmentos, éste se marcaría mediante la formación de una hélice triple con un oligonucleótido fluorescente, antes de la extensión de los fragmentos del DNA sobre el portaobjetos (Geron-Landre et al., 2003). Es importante resaltar que al analizar cada molécula de DNA mediante mapeo óptico se incrementará la precisión de la información posicional, con el consiguiente descenso en el número de clones que deberán ser secuenciados para obtener una secuencia consenso de calidad.

Si bien la automatización todavía requiere mucho trabajo de puesta a punto, el WTSI ha mostrado gran interés por los resultados obtenidos, y ha comenzado, recientemente, a implementar la técnica descrita en esta tesis doctoral.

5.2. El centrómero en *D. melanogaster*

En consistencia con los cambios morfológicos que sufren los cromosomas Y de *Drosophila*, Berloco y colaboradores encontraron, dentro del subgrupo *melanogaster*, que el cromosoma Y es metacéntrico o telocéntrico (figura 33) (Berloco et al., 2005). Esto parece mostrar que, de manera recurrente, han tenido lugar reordenamientos intra-cromosómicos, probablemente mediante inversiones.

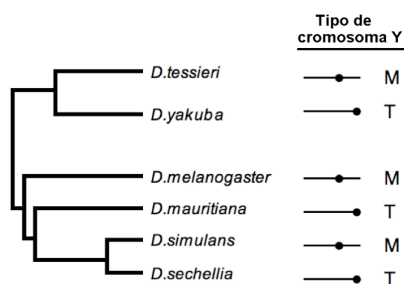


Figura 33. Tipo de cromosoma Y en las especies del subgrupo *melanogaster*, en función de la posición del centrómero.

T indica telocéntrico y M, metacéntrico. Datos extraídos de Berloco et al (2005).

Además, observaron que en la región centromérica, detectada con CID, había secuencias relacionadas con *HeT-A*, independientemente de la posición del centrómero dentro del cromosoma Y. Estos resultados hicieron que Berloco y colaboradores apoyaran la hipótesis según la cual el centrómero del cromosoma Y derivaría de un telómero (Agudo et al., 1999; Villasante et al., 2007a). Finalmente, la obtención de la secuencia del BACR26J21 y el análisis de la misma han proporcionado la información necesaria para mostrar que, efectivamente, la región centromérica h18 de *D. melanogaster* evolucionó a partir de un telómero, muy probablemente mediante una inversión pericéntrica de un cromosoma ancestral telocéntrico.

El aislamiento de varios BACs derivados de la región h17 ha permitido construir un “contig” de unas 500 kb, en una región completamente desconocida hasta el presente trabajo. Este “contig” junto con el “contig” de h18 de más de 600 kb y el de h14 de más de 200 kb (Abad et al., 2004a) representan los únicos clones que han sido aislados del cromosoma Y. Además, la secuencia obtenida de las regiones h18 (160 kb del BACR26J21) y h17 (198 kb del BACR07N15) representan más del 50% de la secuencia del cromosoma Y conocida hasta la fecha. Es más, de toda la secuencia existente del cromosoma Y en las bases de datos, las secuencias generadas en esta tesis doctoral son las únicas que provienen de BACs, y por lo tanto, no son discontinuas.

El análisis de la secuencia del inserto del BACR07N15 ha mostrado que un fragmento de tan sólo 8 kb (comprendiendo las tres duplicaciones segmentales que se han hallado en la región h17) sufrió una amplificación masiva, se le insertaron algunos transposones y generó un palíndromo de al menos 200 kb. Éste es el primer palíndromo de gran tamaño descrito en el genoma de *D. melanogaster*. En el inserto del BACR07N15 hay casi 200 kb de palíndromo, pero los clones obtenidos al hibridar con la sonda Mst35BΨ (secuencia que aparece varias veces a lo largo del palíndromo), muestran que éste es mayor. Se ha estimado que el palíndromo podría abarcar hasta 300 kb, pero se precisaría más información y datos más concluyentes para obtener el tamaño del mismo con exactitud. En cualquier caso, la presencia de un palíndromo de semejante tamaño es indicativo de los eventos que han tenido lugar durante la

evolución del cromosoma Y. En 1939 McClintock estudió el comportamiento de los extremos rotos de los cromosomas de maíz y describió los procesos de “puente-rotura-fusión” (BFB, del inglés “Breakage-Fusión-Bridge”) (McClintock, 1939). En un ciclo de BFB el cromosoma recientemente roto, tras la fusión de las cromátidas hermanas, da lugar a un dicéntrico y al consiguiente puente en la anafase mitótica. Si el puente se rompe, cada núcleo de las células hijas recibe un cromosoma con un extremo roto. Los ciclos de BFB continúan hasta que uno de los extremos rotos es curado mediante la formación de un nuevo telómero funcional, o bien mediante la inactivación de uno de los centrómeros. Como consecuencia del proceso de BFB se produce la formación de palíndromos de tamaño considerable (ver figura 34).

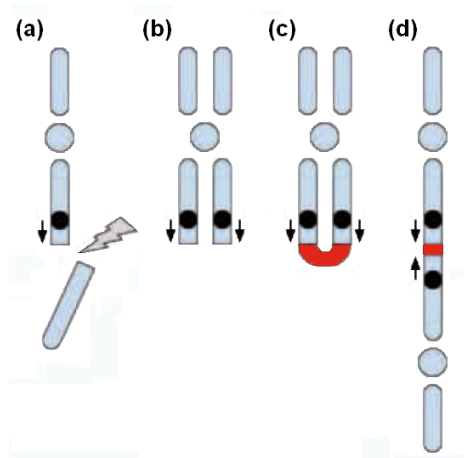


Figura 34. Formación de un gran palíndromo mediante un ciclo de BFB.

(a) Rotura del DNA, (b) replicación y formación de cromátidas hermanas, (c) fusión entre cromátidas y (d) formación de un puente anafásico, y, a su vez, de un palíndromo.

Habiendo observado la presencia del palíndromo en el BACR07N15, cabe pensar que éste fue generado mediante un mecanismo del tipo BFB. Estos procesos de amplificación basados en el mecanismo de “ruptura-fusión-puente” podrían explicar el rápido aumento de tamaño que se ha observado en el cromosoma Y de varias especies de *Drosophila* (Dobzhansky, 1937; Miller y Stone, 1962).

Además, de ser cierto que la región h17 fue generada mediante un ciclo BFB, se esperaría encontrar un centrómero próximo al palíndromo. Ese hecho sugiere que la función centromérica del cromosoma Y de *D. melanogaster* quedaría relegada enteramente a h18.

Resulta especialmente curioso cómo se pudo comenzar el estudio de la región h17, casi fortuitamente, debido a datos publicados en estudios de impronta genética. Así, a partir de la observación de la gran supresión que mostraba un elemento P se ha caracterizado un gran palíndromo en la región h17. Nuestra experiencia demuestra que combinando técnicas y enfoques se pueden conseguir logros y avanzar en el conocimiento de las regiones que presentan mayor dificultad.

El centrómero del cromosoma 3 había sido localizado previamente en la región h53, utilizando reordenamientos cromosómicos (Koryakov et al., 2002). En esta tesis doctoral se ha localizado en la región centromérica PAE del Plato Atlantis de los cromosomas politénicos de mutantes dobles *Su(var)3-9*; *SuUR*. Con más resolución aún, se ha observado que la proteína centromérica CID colocaliza con el satélite centromérico dodeca en fibras de cromatina. Además se puede decir que el estudio de la región centromérica, a nivel molecular ha sido completo pues se ha conseguido pasar del brazo largo al brazo corto del cromosoma 3, cubriendo prácticamente toda la heterocromatina centromérica. Durante el estudio se ha detectado la presencia del satélite 10bp en h52p, mucho más próximo al centrómero de lo que había sido indicado (Lohe et al., 1993). Esta nueva localización se asemeja a la posición que ocupa este satélite en el cromosoma 2. Por el momento se desconoce qué implicación tiene el satélite 10bp en la función centromérica, pero se sabe que la proteína PROD (implicada en la condensación cromosómica en mitosis) se une específicamente a éste satélite (Torok et al., 2000) y que la proteína centromérica MEI-S332, una sugosina, (Lee et al., 2004) localiza en la región cromosómica donde se encuentra el satélite 10bp (Blower y Karpen, 2001). Igualmente, la presencia en el centrómero del cromosoma 3 de secuencias relacionadas con el espaciador intergénico ribosomal (IGSr) ha sido algo inesperado. Además, mediante hibridaciones en cromosomas politénicos de

mutantes dobles *Su(var)3-9*; *SuUR*, se ha confirmado que estas secuencias se encuentran en dos subregiones de PAE siendo una de ellas colindante con el satélite 10bp. Probablemente las secuencias IGSr llegaron al centrómero del cromosoma 3 mediante una duplicación segmental procedente del DNA ribosómico del cromosoma X, o del Y, y posteriormente tuvo lugar la inversión que dio lugar a estas dos regiones. Se desconoce si estas secuencias cumplen alguna función centromérica.

El estudio de los satélites dodeca, 10bp y 15bp en especies del subgrupo *melanogaster* (*D. melanogaster*, *D. simulans* y *D. mauritiana*) muestra la rápida evolución a la que están sometidos los satélites centroméricos. Según los datos obtenidos, parece como si el ancestro común hubiese tenido dodeca presente en el centrómero de los cromosomas 2 y 3, y que en el cromosoma 2 de *D. melanogaster* se hubiese perdido. Esto mismo podría haber sucedido con el satélite 15bp, que está presente en los cromosomas 2 y 3 tanto de *D. simulans* como de *D. mauritiana*, pero no se ha detectado citológicamente en *D. melanogaster*, aunque sí mediante secuenciación, lo cual es indicativo de la pequeña cantidad en que se encuentra. Por otra parte, parece muy probable que el satélite 10bp haya derivado del 15bp, dado que ambos tienen alto contenido en AT. Es más, ambos contienen la secuencia “AACAT”, y el resto de la unidad de repetición del 10bp (AGAAT) es igual, salvo por un nucleótido, a la secuencia complementaria reversa (AGAAC) de las siguientes 5 bases del satélite 15bp. Esta hipótesis concordaría con la visión de un ancestro común, que tendría los satélites dodeca y 15bp en los cromosomas 2 y 3, y mientras que en *D. simulans* y *D. mauritiana* se han mantenido, en *D. melanogaster*, el satélite 15bp habría dado lugar al 10bp, y éste último habría ido reemplazándolo.

Respecto al rastreo de las genotecas en busca de los clones que tuviesen secuencias de dodeca, se ha observado que sólo se obtuvieron dos clones cuyos dos extremos contenían secuencias de dodeca. En un principio se pensó que estos clones pertenecían al interior de uno de los dos bloques grandes de dodeca, presentes en h53. Sin embargo, se ha comprobado que estos clones llevan transposones insertados, por lo que se deduce que no pertenecen a la

zona interna de ninguno de los bloques, sino, probablemente, a la salida de alguno de los mismos. Eso lleva a plantearse el por qué no hay ningún clon cuyo inserto sea únicamente secuencias del satélite dodeca. Podría ser que este satélite no puede ser clonado, o bien que es altamente inestable, y que una vez clonado tiende a delecionarse. Esta segunda hipótesis ha sido confirmada en el presente trabajo, a través del BACR29J09. Este BAC pertenece a una genoteca que se generó clonando fragmentos de 100 kb, obtenidos mediante DNA de moscas cizallado mecánicamente. Sin embargo, al secuenciar el clon sólo se ha obtenido una secuencia de 23 kb, que incluye unas 20 kb de secuencias IGSr, y tan solo 3 kb de dodeca. Es más, los extremos de este BAC habían sido previamente secuenciados, y la secuencia del extremo que lleva dodeca no ha sido encontrada en el ensamblado CP000331, que tiene un tamaño de 60 kb. Resulta evidente que alrededor de 80 kb de dodeca estaban inicialmente presentes en el BAC, y que se han delecionado del mismo. Lamentablemente, cuanto mayor es el contenido en dodeca en un BAC, más inestable es éste, y aquellos BACs que tienen bajo contenido no resultan de tanto interés, pues se alejan de la región centromérica.

Para concluir es necesario preguntarse qué tienen en común los dos centrómeros estudiados. Como se ha visto, el centrómero del cromosoma Y contiene el satélite 18HT y éste está, a su vez, formado por fragmentos de retrotransposones teloméricos. Por su parte, el centrómero del cromosoma 3 está compuesto, fundamentalmente, por el satélite dodeca. Es evidente que no existe una secuencia única que esté presente en todos los centrómeros, ni entre especies ni tan siquiera dentro de una misma especie, como se ha observado en esta tesis doctoral. A pesar de esto, es posible que exista un determinante estructural en el DNA centromérico, independiente de la secuencia primaria, que al ser reconocido por las proteínas centroméricas sirva para nuclear la formación del cinetocoro. *A priori* no se observa parecido alguno entre el satélite dodeca, de 11bp y 12bp de unidad de repetición, y el satélite 18HT, de 3,1 kb y 2 kb de unidad de repetición, formado a su vez por las regiones 3'UTR de 3 elementos *HeT-A*, además de un fragmento derivado de *TART*. No obstante, se ha visto que tanto el satélite dodeca como la región

3'UTR del elemento *HeT-A* tienen una distribución asimétrica de guaninas y citosinas, es decir, una de las cadenas del DNA es más rica en Gs (Abad et al., 1992). Además, los análisis estructurales han mostrado, por una parte, que el satélite dodeca es capaz de formar estructuras plegadas que se estabilizan por apareamientos no Watson-Crick, tipo G2 (Ferrer et al., 1995) y G6 (Chou y Chin, 2001) y por otra parte, que la región 3'UTR de los elementos *HeT-A* tiene secuencias propensas a formar estructuras plegadas del tipo G4 DNA (Abad y Villasante, 1999). Igualmente, las secuencias añadidas por la telomerasa tienen la capacidad de formar G4 DNA *in vitro* (Rhodes y Giraldo, 1995) e *in vivo* (Paeschke et al., 2005).

En definitiva, parece que la secuencia primaria no es determinante para la función centromérica, y esto explicaría la diversidad tan elevada de los satélites centroméricos. Sin embargo, sí que podría existir un requerimiento de composición de la secuencia para formar estructuras secundarias de DNA, que promoverían la formación del cinetocoro, por un mecanismo aún desconocido. Si esto fuese así, parece que tanto para la función centromérica como para la función telomérica se requerirían estructuras secundarias de DNA similares. Esta reflexión, junto con la conocida capacidad de los telómeros de interaccionar con los microtúbulos en meiosis para promover el movimiento de los cromosomas (Goday y Pimpinelli, 1989; Manzanero y Puertas, 2003; Östergren y Prakken, 1946; Perez et al., 1997; Rhoades y Vilkomerson, 1942), nos ha llevado a proponer que los centrómeros podrían haber derivado de los telómeros en el origen del cromosoma eucariótico (ver figura 35) (Villasante et al., 2007a; Villasante et al., 2007c).

Dos estudios recientes sobre la función centromérica en mitosis han dado lugar a resultados que están en concordancia con la hipótesis propuesta. Takahashi y colaboradores han visto que, tras eliminar el centrómero de un cromosoma en *Schizosaccharomyces pombe*, se formaban neocentrómeros en las regiones subteloméricas (Ishii et al., 2008). De forma similar, Heun y colaboradores han observado en *D. melanogaster* que, tras sobreexpresar la proteína centromérica CID, se formaban neocentrómeros en las regiones teloméricas (Heun, comunicación personal).

La prospectiva del trabajo presentado en esta tesis doctoral implicaría realizar ensayos funcionales con alguna de las secuencias centroméricas descritas. Sería de interés especial realizarlo empleando secuencias del satélite 18HT, pues si se formase un centrómero *de novo*, se obtendría, además, una evidencia a favor de que una secuencia telomérica pueda formar un centrómero funcional.

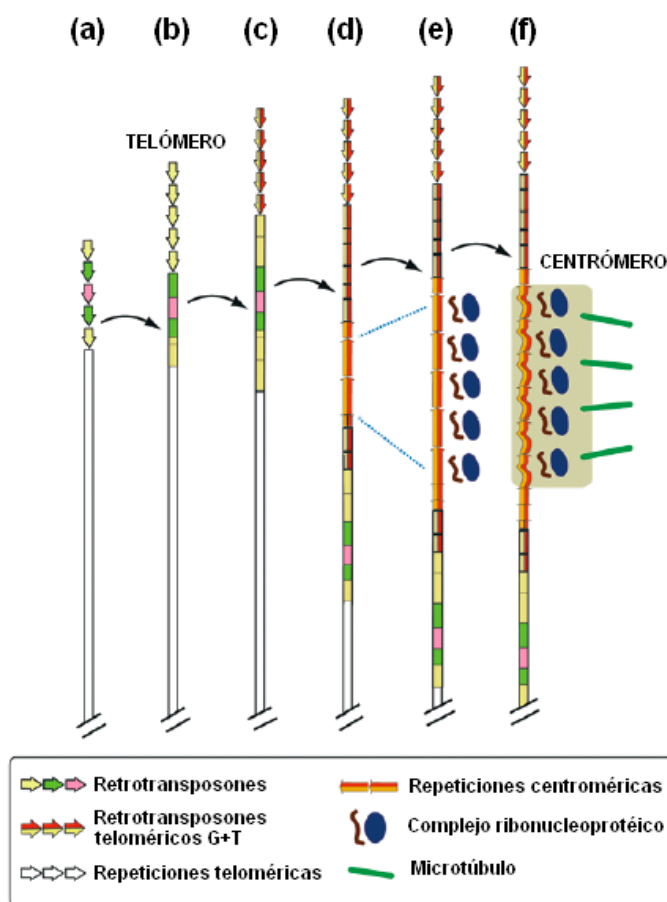


Figura 35. Modelo sobre el origen del centrómero.

(a) Los retroelementos fueron movilizados para curar los extremos de los primeros cromosomas lineales. (b) Un retrotransposón específico de telómeros fue seleccionado. (c) El retrotransposón más eficiente para generar el "capping" fue posteriormente seleccionado. (d) Aparecieron diferentes repeticiones subtroméricas. (e) Una repetición subtromérica fue amplificada. La transcripción de estas repeticiones promovió la formación de complejos ribonucleoprotéicos. (f) Esta región subtromérica dio lugar a un proto-centrómero tras ser reconocida por los microtúbulos. Extraído de Villasante et al (2007b).

Conclusiones

6. CONCLUSIONES

1. Se ha desarrollado una técnica de secuenciación mediante transposición que, empleando la posición de inserción del transposón durante el proceso de ensamblaje, permite obtener la secuencia de DNA altamente repetido clonado en cromosomas artificiales de bacterias,
2. El análisis de la secuencia de un clon heterocromático muestra que la región centromérica h18 derivó de un telómero ancestral.
3. Se han obtenido los primeros clones y secuencia de la región h17 del cromosoma Y de *D. melanogaster*. Gracias a ello, se ha detectado la presencia de un palíndromo de más de 200 kb, probablemente generado mediante un mecanismo de “ruptura-fusión-puente”.
4. El satélite dodeca, específico de la región centromérica h53 del cromosoma 3 de *D. melanogaster*, colocaliza con la proteína centromérica CID en fibras de cromatina.
5. Se han caracterizado los satélites de DNA, los elementos transponibles y las duplicaciones segmentales que se hallan en la región centromérica del cromosoma 3. El análisis realizado cubre el hueco que existía en la heterocromatina de este cromosoma.

Bibliografía

7. BIBLIOGRAFÍA

- Abad, J.P., M. Carmena, S. Baars, R.D. Saunders, D.M. Glover, P. Ludena, C. Sentis, C. Tyler-Smith y A. Villasante. 1992. Dodeca satellite: a conserved G+C-rich satellite from the centromeric heterochromatin of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 89:4663-7.
- Abad, J.P., B. de Pablos, M. Agudo, I. Molina, G. Giovinnazzo, A. Martin-Gallardo y A. Villasante. 2004a. Genomic and cytological analysis of the Y chromosome of *Drosophila melanogaster*: telomere-derived sequences at internal regions. *Chromosoma*. 113:295-304.
- Abad, J.P., B. De Pablos, K. Osoegawa, P.J. De Jong, A. Martin-Gallardo y A. Villasante. 2004b. Genomic analysis of *Drosophila melanogaster* telomeres: full-length copies of HeT-A and TART elements at telomeres. *Mol Biol Evol*. 21:1613-9.
- Abad, J.P., B. De Pablos, K. Osoegawa, P.J. De Jong, A. Martin-Gallardo y A. Villasante. 2004c. TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Mol Biol Evol*. 21:1620-4.
- Abad, J.P. y A. Villasante. 1999. The 3' non-coding region of the *Drosophila melanogaster* HeT-A telomeric retrotransposon contains sequences with propensity to form G-quadruplex DNA. *FEBS Lett*. 453:59-62.
- Agudo, M., A. Losada, J.P. Abad, S. Pimpinelli, P. Ripoll y A. Villasante. 1999. Centromeres from telomeres? The centromeric region of the Y chromosome of *Drosophila melanogaster* contains a tandem array of telomeric HeT-A- and TART-related sequences. *Nucleic Acids Res*. 27:3318-24.
- Allshire, R.C., E.R. Nimmo, K. Ekwall, J.P. Javerzat y G. Cranston. 1995. Mutations derepressing silent centromeric domains in fission yeast disrupt chromosome segregation. *Genes Dev*. 9:218-33.
- Amor, D.J., K. Bentley, J. Ryan, J. Perry, L. Wong, H. Slater y K.H. Choo. 2004. Human centromere repositioning "in progress". *Proc Natl Acad Sci U S A*. 101:6542-7.
- Andreyeva, E.N., T.D. Kolesnikova, O.V. Demakova, M. Mendez-Lago, G.V. Pokholkova, E.S. Belyaeva, F. Rossi, P. Dimitri, A. Villasante y I.F. Zhimulev. 2007. High-resolution analysis of *Drosophila* heterochromatin organization using SuUR Su(var)3-9 double mutants. *Proc Natl Acad Sci U S A*. 104:12819-24.
- Argast, G.M., K.M. Stephens, M.J. Emond y R.J. Monnat, Jr. 1998. I-Ppol and I-Crel homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J Mol Biol*. 280:345-53.
- Berlaco, M., L. Fanti, F. Sheen, R.W. Levis y S. Pimpinelli. 2005. Heterochromatic distribution of HeT-A- and TART-like sequences in several *Drosophila* species. *Cytogenet Genome Res*. 110:124-33.
- Bernard, P., J.F. Maure, J.F. Partridge, S. Genier, J.P. Javerzat y R.C. Allshire. 2001. Requirement of heterochromatin for cohesion at centromeres. *Science*. 294:2539-42.
- Blackburn, E.H. 1994. Telomeres: no end in sight. *Cell*. 77:621-3.

- Blower, M.D. y G.H. Karpen. 2001. The role of *Drosophila* CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nat Cell Biol.* 3:730-9.
- Blower, M.D., B.A. Sullivan y G.H. Karpen. 2002. Conserved organization of centromeric chromatin in flies and humans. *Dev Cell.* 2:319-30.
- Bonfield, J.K., K. Smith y R. Staden. 1995. A new DNA sequence assembly program. *Nucleic Acids Res.* 23:4992-9.
- Brennecke, J., A.A. Aravin, A. Stark, M. Dus, M. Kellis, R. Sachidanandam y G.J. Hannon. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell.* 128:1089-103.
- Bridges, C.B. 1935. Salivary chromosome maps with a key to the banding of the chromosomes of *Drosophila melanogaster*. *J Heredity.* 26:60-64.
- Brown, S.W. 1966. Heterochromatin. *Science.* 151:417-25.
- Carmena, M., J.P. Abad, A. Villasante y C. Gonzalez. 1993. The *Drosophila melanogaster* dodecasatellite sequence is closely linked to the centromere and can form connections between sister chromatids during mitosis. *J Cell Sci.* 105 (Pt 1):41-50.
- Carmena, M. y C. Gonzalez. 1995. Transposable elements map in a conserved pattern of distribution extending from beta-heterochromatin to centromeres in *Drosophila melanogaster*. *Chromosoma.* 103:676-84.
- Carvalho, A.B., B.A. Dobo, M.D. Vrbancovskiy y A.G. Clark. 2001. Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A.* 98:13225-30.
- Choo, K.H., B. Vissel, A. Nagy, E. Earle y P. Kalitsis. 1991. A survey of the genomic distribution of alpha satellite DNA on all the human chromosomes, and derivation of a new consensus sequence. *Nucleic Acids Res.* 19:1179-82.
- Chou, S.H. y K.H. Chin. 2001. Quadruple intercalated G-6 stack: a possible motif in the fold-back structure of the *Drosophila* centromeric dodecasatellite? *J Mol Biol.* 314:139-52.
- Clarke, L. 1990. Centromeres of budding and fission yeasts. *Trends Genet.* 6:150-4.
- Clarke, L., H. Amstutz, B. Fishel y J. Carbon. 1986. Analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. *Proc Natl Acad Sci U S A.* 83:8253-7.
- Clarke, L. y J. Carbon. 1980. Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature.* 287:504-9.
- Clarke, L. y J. Carbon. 1985. The structure and function of yeast centromeres. *Annu Rev Genet.* 19:29-56.
- Consortium, R.G.S.P. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature.* 428:493-521.
- Cryderman, D.E., E.J. Morris, H. Biessmann, S.C. Elgin y L.L. Wallrath. 1999. Silencing at *Drosophila* telomeres: nuclear organization and chromatin structure play critical roles. *Embo J.* 18:3724-35.
- Danilevskaya, O., A. Lofsky, E.V. Kurenova y M.L. Pardue. 1993. The Y chromosome of *Drosophila melanogaster* contains a distinctive subclass of Het-A-related repeats. *Genetics.* 134:531-43.
- Danilevskaya, O., F. Slot, M. Pavlova y M.L. Pardue. 1994. Structure of the *Drosophila* HeT-A transposon: a retrotransposon-like element forming telomeres. *Chromosoma.* 103:215-24.

- Danilevskaya, O.N., I.R. Arkhipova, K.L. Traverse y M.L. Pardue. 1997. Promoting in tandem: the promoter for telomere transposon HeT-A and implications for the evolution of retroviral LTRs. *Cell*. 88:647-55.
- Danilevskaya, O.N., D.A. Petrov, M.N. Pavlova, A. Koga, E.V. Kurenova y D.L. Hartl. 1992. A repetitive DNA element, associated with telomeric sequences in *Drosophila melanogaster*, contains open reading frames. *Chromosoma*. 102:32-40.
- de Lange, T. 1992. Human telomeres are attached to the nuclear matrix. *Embo J*. 11:717-24.
- de Lange, T. 1994. Activation of telomerase in a human tumor. *Proc Natl Acad Sci U S A*. 91:2882-5.
- De Lange, T., J.M. Kooter, P.A. Michels y P. Borst. 1983. Telomere conversion in trypanosomes. *Nucleic Acids Res*. 11:8149-65.
- Dernburg, A.F., K.W. Broman, J.C. Fung, W.F. Marshall, J. Philips, D.A. Agard y J.W. Sedat. 1996a. Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell*. 85:745-59.
- Dernburg, A.F., J.W. Sedat y R.S. Hawley. 1996b. Direct evidence of a role for heterochromatin in meiotic chromosome segregation. *Cell*. 86:135-46.
- Dobzhansky, T. 1937. Further Data on the Variation of the Y Chromosome in *Drosophila Pseudoobscura*. *Genetics*. 22:340-6.
- Earnshaw, W.C. y N. Rothfield. 1985. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*. 91:313-21.
- Earnshaw, W.C., K.F. Sullivan, P.S. Machlin, C.A. Cooke, D.A. Kaiser, T.D. Pollard, N.F. Rothfield y D.W. Cleveland. 1987. Molecular cloning of cDNA for CENP-B, the major human centromere autoantigen. *J Cell Biol*. 104:817-29.
- Ekwall, K. 2007. Epigenetic control of centromere behavior. *Annu Rev Genet*. 41:63-81.
- Elgin, S.C. y S.I. Grewal. 2003. Heterochromatin: silence is golden. *Curr Biol*. 13:R895-8.
- Fang, G. y T.R. Cech. 1993. The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell*. 74:875-85.
- Ferrer, N., F. Azorin, A. Villasante, C. Gutierrez y J.P. Abad. 1995. Centromeric dodeca-satellite DNA sequences form fold-back structures. *J Mol Biol*. 245:8-21.
- Gatti, M. y S. Pimpinelli. 1992. Functional elements in *Drosophila melanogaster* heterochromatin. *Annu Rev Genet*. 26:239-75.
- George, J.A., P.G. DeBaryshe, K.L. Traverse, S.E. Celniker y M.L. Pardue. 2006. Genomic organization of the *Drosophila* telomere retrotransposable elements. *Genome Res*. 16:1231-40.
- Geron-Landre, B., T. Roulon, P. Desbiolles y C. Escude. 2003. Sequence-specific fluorescent labeling of double-stranded DNA observed at the single molecule level. *Nucleic Acids Res*. 31:e125.
- Giacalone, J., S. Delobette, V. Gibaja, L. Ni, Y. Skiadas, R. Qi, J. Edington, Z. Lai, D. Gebauer, H. Zhao, T. Anantharaman, B. Mishra, L.G. Brown, R. Saxena, D.C. Page y D.C. Schwartz. 2000. Optical mapping of BAC clones from the human Y chromosome DAZ locus. *Genome Res*. 10:1421-9.

- Gilson, E., T. Laroche y S.M. Gasser. 1993. Telomeres and the functional architecture of the nucleus. *Trends Cell Biol.* 3:128-34.
- Gimble, F.S. y J. Thorner. 1993. Purification and characterization of VDE, a site-specific endonuclease from the yeast *Saccharomyces cerevisiae*. *J Biol Chem.* 268:21844-53.
- Gimble, F.S. y J. Wang. 1996. Substrate recognition and induced DNA distortion by the PI-SceI endonuclease, an enzyme generated by protein splicing. *J Mol Biol.* 263:163-80.
- Goday, C. y S. Pimpinelli. 1989. Centromere organization in meiotic chromosomes of *Parascaris univalens*. *Chromosoma.* 98:160-6.
- Goshima, G., T. Kiyomitsu, K. Yoda y M. Yanagida. 2003. Human centromere chromatin protein hMis12, essential for equal segregation, is independent of CENP-A loading pathway. *J Cell Biol.* 160:25-39.
- Gottschling, D.E., O.M. Aparicio, B.L. Billington y V.A. Zakian. 1990. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell.* 63:751-62.
- Grewal, S.I. y S. Jia. 2007. Heterochromatin revisited. *Nat Rev Genet.* 8:35-46.
- Gvozdev, V.A., G.L. Kogan y L.A. Usakin. 2005. The Y chromosome as a target for acquired and amplified genetic material in evolution. *Bioessays.* 27:1256-62.
- Hall, I.M., G.D. Shankaranarayana, K. Noma, N. Ayoub, A. Cohen y S.I. Grewal. 2002. Establishment and maintenance of a heterochromatin domain. *Science.* 297:2232-7.
- Hall, S.E., G. Kettler y D. Preuss. 2003. Centromere satellites from Arabidopsis populations: maintenance of conserved and variable domains. *Genome Res.* 13:195-205.
- Harley, C.B. y B. Villeponteau. 1995. Telomeres and telomerase in aging and cancer. *Curr Opin Genet Dev.* 5:249-55.
- Harrington, J.J., G. Van Bokkelen, R.W. Mays, K. Gustashaw y H.F. Willard. 1997. Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nat Genet.* 15:345-55.
- Heitz, E. 1928. Das heterochromatin der Moose. *I. Jahr Wiss Bot.* 69:762-818.
- Henikoff, S. 1990. Position-effect variegation after 60 years. *Trends Genet.* 6:422-6.
- Hoskins, R.A., J.W. Carlson, C. Kennedy, D. Acevedo, M. Evans-Holm, E. Frise, K.H. Wan, S. Park, M. Mendez-Lago, F. Rossi, A. Villasante, P. Dimitri, G.H. Karpen y S.E. Celniker. 2007. Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science.* 316:1625-8.
- Hoskins, R.A., C.R. Nelson, B.P. Berman, T.R. Lavery, R.A. George, L. Ciesiolka, M. Naeemuddin, A.D. Arenson, J. Durbin, R.G. David, P.E. Tabor, M.R. Bailey, D.R. DeShazo, J. Catanese, A. Mammoser, K. Osoegawa, P.J. de Jong, S.E. Celniker, R.A. Gibbs, G.M. Rubin y S.E. Scherer. 2000. A BAC-based physical map of the major autosomes of *Drosophila melanogaster*. *Science.* 287:2271-4.
- Howman, E.V., K.J. Fowler, A.J. Newson, S. Redward, A.C. MacDonald, P. Kalitsis y K.H. Choo. 2000. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. *Proc Natl Acad Sci U S A.* 97:1148-53.

- Ishii, K., Y. Ogiyama, Y. Chikashige, S. Soejima, F. Masuda, T. Kakuma, Y. Hiraoka y K. Takahashi. 2008. Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. *Science*. 321:1088-91.
- Jasin, M. 1996. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* 12:224-8.
- Jenuwein, T. y C.D. Allis. 2001. Translating the histone code. *Science*. 293:1074-80.
- Kaminker, J.S., C.M. Bergman, B. Kronmiller, J. Carlson, R. Svirskas, S. Patel, E. Frise, D.A. Wheeler, S.E. Lewis, G.M. Rubin, M. Ashburner y S.E. Celniker. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3:RESEARCH0084.
- Kang, Y., T. Durfee, J.D. Glasner, Y. Qiu, D. Frisch, K.M. Winterberg y F.R. Blattner. 2004. Systematic mutagenesis of the *Escherichia coli* genome. *J Bacteriol.* 186:4921-30.
- Karpen, G.H. y R.C. Allshire. 1997. The case for epigenetic effects on centromere identity and function. *Trends Genet.* 13:489-96.
- Karpen, G.H., M.H. Le y H. Le. 1996. Centric heterochromatin and the efficiency of achiasmate disjunction in *Drosophila* female meiosis. *Science*. 273:118-22.
- Kellum, R. y B.M. Alberts. 1995. Heterochromatin protein 1 is required for correct chromosome segregation in *Drosophila* embryos. *J. Cell Sci.* 108:1419-31.
- Koryakov, D.E., I.F. Zhimulev y P. Dimitri. 2002. Cytogenetic analysis of the third chromosome heterochromatin of *Drosophila melanogaster*. *Genetics*. 160:509-17.
- Kouzarides, T. 2007. Chromatin modifications and their function. *Cell*. 128:693-705.
- Lee, J.Y., K.J. Dej, J.M. Lopez y T.L. Orr-Weaver. 2004. Control of centromere localization of the MEI-S332 cohesion protection protein. *Curr Biol*. 14:1277-83.
- Levis, R.W., R. Ganesan, K. Houtchens, L.A. Tolar y F.M. Sheen. 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell*. 75:1083-93.
- Lippman, Z. y R. Martienssen. 2004. The role of RNA interference in heterochromatic silencing. *Nature*. 431:364-70.
- Lohe, A.R., A.J. Hilliker y P.A. Roberts. 1993. Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics*. 134:1149-74.
- Lohe, A.R. y P.A. Roberts. 1988. Evolution of satellite DNA sequences in *Drosophila*, en "Heterochromatin". Editado por Ram S. Verma. Cambridge University Press.
- Losada, A., J.P. Abad, M. Agudo y A. Villasante. 1999a. The analysis of Circe, an LTR retrotransposon of *Drosophila melanogaster*, suggests that an insertion of non-LTR retrotransposons into LTR elements can create chimeric retroelements. *Mol Biol Evol.* 16:1341-6.
- Losada, A., J.P. Abad, M. Agudo y A. Villasante. 2000. Long-range analysis of the centromeric region of *Drosophila melanogaster* chromosome 3. *Chromosome Res.* 8:651-3.

- Losada, A., J.P. Abad y A. Villasante. 1997. Organization of DNA sequences near the centromere of the *Drosophila melanogaster* Y chromosome. *Chromosoma*. 106:503-12.
- Losada, A., M. Agudo, J.P. Abad y A. Villasante. 1999b. HeT-A telomere-specific retrotransposons in the centric heterochromatin of *Drosophila melanogaster* chromosome 3. *Mol Gen Genet*. 262:618-22.
- Louis, E.J. 1995. The chromosome ends of *Saccharomyces cerevisiae*. *Yeast*. 11:1553-73.
- Lyon, M.F. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*. 190:372-3.
- Maddox, P.S., K. Oegema, A. Desai y I.M. Cheeseman. 2004. "Holo"er than thou: chromosome segregation and kinetochore function in *C. elegans*. *Chromosome Res*. 12:641-53.
- Maggert, K.A. y K.G. Golic. 2002. The Y chromosome of *Drosophila melanogaster* exhibits chromosome-wide imprinting. *Genetics*. 162:1245-58.
- Maio, J.J. 1971. DNA strand reassociation and polyribonucleotide binding in the African green monkey, *Cercopithecus aethiops*. *J Mol Biol*. 56:579-95.
- Makalowski, W. 2003. Genomics. Not junk after all. *Science*. 300:1246-7.
- Manzanero, S. y M.J. Puertas. 2003. Rye terminal neocentromeres: characterisation of the underlying DNA and chromatin structure. *Chromosoma*. 111:408-15.
- Mason, J.M. y H. Biessman. 1995. The unusual telomeres of *Drosophila*. *Trends Genet*. 11:58-62.
- Masumoto, H., H. Masukata, Y. Muro, N. Nozaki y T. Okazaki. 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol*. 109:1963-73.
- McClintock, B. 1939. The Behavior in Successive Nuclear Divisions of a Chromosome Broken at Meiosis. *Proc Natl Acad Sci U S A*. 25:405-16.
- McCord, R.A. y D. Broccoli. 2008. Telomeric chromatin: Roles in aging, cancer and hereditary disease. *Mutat Res*.
- McKee, B.D. y G.H. Karpen. 1990. *Drosophila* ribosomal RNA genes function as an X-Y pairing site during male meiosis. *Cell*. 61:61-72.
- Meyne, J., R.L. Ratliff y R.K. Moyzis. 1989. Conservation of the human telomere sequence (TTAGGG)_n among vertebrates. *Proc Natl Acad Sci U S A*. 86:7049-53.
- Miller, D.D. y L.E. Stone. 1962. Reinvestigation of karyotype in *Drosophila affinis* Sturtevant and related species. *J Hered*. 53:12-24.
- Mitchel, A.R., J.R. Godsen y D.A. Miller. 1985. A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma*. 92:369-377.
- Monteilhet, C., A. Perrin, A. Thierry, L. Colleaux y B. Dujon. 1990. Purification and characterization of the in vitro activity of I-Sce I, a novel and highly specific endonuclease encoded by a group I intron. *Nucleic Acids Res*. 18:1407-13.
- Muller, H.J. 1938. The remaking of chromosomes. *The collecting Net- Woods Hole*. 13:181-195.
- Muller, H.J. y E. Altenburg. 1930. The Frequency of Translocations Produced by X-Rays in *Drosophila*. *Genetics*. 15:283-311.

- Murphy, T.D. y G.H. Karpen. 1995. Localization of centromere function in a *Drosophila* minichromosome. *Cell*. 82:599-609.
- Myers, E.W., G.G. Sutton, A.L. Delcher, I.M. Dew, D.P. Fasulo, M.J. Flanigan, S.A. Kravitz, C.M. Mobarry, K.H. Reinert, K.A. Remington, E.L. Anson, R.A. Bolanos, H.H. Chou, C.M. Jordan, A.L. Halpern, S. Lonardi, E.M. Beasley, R.C. Brandon, L. Chen, P.J. Dunn, Z. Lai, Y. Liang, D.R. Nusskern, M. Zhan, Q. Zhang, X. Zheng, G.M. Rubin, M.D. Adams y J.C. Venter. 2000. A whole-genome assembly of *Drosophila*. *Science*. 287:2196-204.
- Ohno, S. 1972. So much "junk" DNA in our genome. *Brookhaven Symp Biol*. 23:366-70.
- Östergren, G. y R. Prakken. 1946. Behaviour on the spindle of actively mobile chromosome ends of rye. *Hereditas*. 32.
- Paeschke, K., T. Simonsson, J. Postberg, D. Rhodes y H.J. Lipps. 2005. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol*. 12:847-54.
- Pardue, M.L., O.N. Danilevskaya, K. Lowenhaupt, F. Slot y K.L. Traverse. 1996. *Drosophila* telomeres: new views on chromosome evolution. *Trends Genet*. 12:48-52.
- Pardue, M.L., O.N. Danilevskaya, K.L. Traverse y K. Lowenhaupt. 1997. Evolutionary links between telomeres and transposable elements. *Genetica*. 100:73-84.
- Perez, R., F. Panzera, J. Page, J.A. Suja y J.S. Rufas. 1997. Meiotic behaviour of holocentric chromosomes: orientation and segregation of autosomes in *Triatoma infestans* (Heteroptera). *Chromosome Res*. 5:47-56.
- Peters, A.H., D. O'Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schofer, K. Weipoltshammer, M. Pagani, M. Lachner, A. Kohlmaier, S. Opravil, M. Doyle, M. Sibilia y T. Jenuwein. 2001a. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*. 107:323-37.
- Peters, A.H., D. O'Carroll, H. Scherthan, K. Mechtler, S. Sauer, C. Schöfer, K. Weipoltshammer, M. Pagani, M. Lachner, A. Kohlmaier, S. Opravil, M. Doyle, M. Sibilia y T. Jenuwein. 2001b. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell*. 107:323-37.
- Pimpinelli, S., M. Berloco, L. Fanti, P. Dimitri, S. Bonacorssi, E. Marchetti, R. Caizzi, C. Caggese y M. Gatti. 1994. Transposable elements are structural components of *Drosophila melanogaster* heterochromatin. *Proc Natl Acad Sci U S A*. 92:3804-3808.
- Pirrotta, V., C. Hadfield y G.H.J. Pretorius. 1983. Microdissection and cloning of the white locus and the 3B1-3C2 region of the *Drosophila* X chromosome. *The EMBO Journal*.
- Pluta, A.F., A.M. Mackay, A.M. Ainsztein, I.G. Goldberg y W.C. Earnshaw. 1995. The centromere: hub of chromosomal activities. *Science*. 270:1591-4.
- Prades, C., A.M. Laurent, J. Puechberty, Y. Yurov y G. Roizes. 1996. SINE and LINE within human centromeres. *J Mol Evol*. 42:37-43.
- Ren, X., L. Eisenhour, C. Hong, Y. Lee y B.D. McKee. 1997. Roles of rDNA spacer and transcription unit-sequences in X-Y meiotic chromosome pairing in *Drosophila melanogaster* males. *Chromosoma*. 106:29-36.

- Rhoades, M.M. y H. Vilkomerson. 1942. On the Anaphase Movement of Chromosomes. *Proc Natl Acad Sci U S A*. 28:433-6.
- Rhodes, D. y R. Giraldo. 1995. Telomere structure and function. *Curr Opin Struct Biol*. 5:311-22.
- Riethman, H. 2008. Human telomere structure and biology. *Annu Rev Genomics Hum Genet*. 9:1-19.
- Rockmill, B. y G.S. Roeder. 1998. Telomere-mediated chromosome pairing during meiosis in budding yeast. *Genes Dev*. 12:2574-86.
- Rudd, M.K. y H.F. Willard. 2004. Analysis of the centromeric regions of the human genome assembly. *Trends Genet*. 20:529-33.
- Sambrook, J., E.F. Fritsch y T. Maniatis. 1989. Molecular cloning: a laboratory manual. *Cold Spring Harbor Laboratory press, Cold Spring Harbor, NY*.
- Schueler, M.G., A.W. Higgins, M.K. Rudd, K. Gustashaw y H.F. Willard. 2001. Genomic and genetic definition of a functional human centromere. *Science*. 294:109-15.
- Schultz, J. 1936. Variegation in *Drosophila* and the Inert Chromosome Regions. *Proc Natl Acad Sci U S A*. 22:27-33.
- Sen, D. y W. Gilbert. 1988. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*. 334:364-6.
- Sheen, F.M. y R.W. Levis. 1994. Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. *Proc Natl Acad Sci U S A*. 91:12510-4.
- Smit, A.F. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev*. 9:657-63.
- Smith, C.D., S. Shu, C.J. Mungall y G.H. Karpen. 2007. The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science*. 316:1586-91.
- Southern, E.M., R. Anand, W.R.A. Brown y D.S. Fletcher. 1987. A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res*. 15:5925-5943.
- Staden, R. 1996. The Staden sequence analysis package. *Mol Biotechnol*. 5:233-41.
- Sullivan, B.A. y G.H. Karpen. 2004. Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol*. 11:1076-83.
- Sun, X., J. Wahlstrom y G. Karpen. 1997. Molecular structure of a functional *Drosophila* centromere. *Cell*. 91:1007-19.
- Sundquist, W.I. y A. Klug. 1989. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*. 342:825-9.
- Topp, C.N., C.X. Zhong y R.K. Dawe. 2004. Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci U S A*. 101:15986-91.
- Torok, T., M. Gorjanacz, P.J. Bryant y I. Kiss. 2000. Prod is a novel DNA-binding protein that binds to the 1.686 g/cm(3) 10 bp satellite repeat of *Drosophila melanogaster*. *Nucleic Acids Res*. 28:3551-7.
- Traverse, K.L. y M.L. Pardue. 1989. Studies of He-T DNA sequences in the pericentric regions of *Drosophila* chromosomes. *Chromosoma*. 97:261-71.

- Vagnarelli, P., S.A. Ribeiro y W.C. Earnshaw. 2008. Centromeres: old tales and new tools. *FEBS Lett.* 582:1950-9.
- Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huson, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A.J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neelam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al. 2001. The sequence of the human genome. *Science.* 291:1304-51.
- Vibranovski, M.D., L.B. Koerich y A.B. Carvalho. 2008. Two new Y-linked genes in *Drosophila melanogaster*. *Genetics.* 179:2325-7.
- Villasante, A., J.P. Abad y M. Mendez-Lago. 2007a. Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc Natl Acad Sci U S A.* 104:10542-7.
- Villasante, A., J.P. Abad, R. Planello, M. Mendez-Lago, S.E. Celniker y B. de Pablos. 2007b. *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res.* 17:1909-18.
- Villasante, A., M. Mendez-Lago, J.P. Abad y E. Montejo de Garcini. 2007c. The birth of the centromere. *Cell Cycle.* 6:2872-6.
- Volpe, T.A., C. Kidner, I.M. Hall, G. Teng, S.I. Grewal y R.A. Martienssen. 2002. Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science.* 297:1833-7.
- Voullaire, L.E., H.R. Slater, V. Petrovic y K.H. Choo. 1993. A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere? *Am J Hum Genet.* 52:1153-63.
- Wei, W., N. Gilbert, S.L. Ooi, J.F. Lawler, E.M. Ostertag, H.H. Kazazian, J.D. Boeke y J.V. Moran. 2001. Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol.* 21:1429-39.
- Weiler, K.S. y B.T. Wakimoto. 1995. Heterochromatin and gene expression in *Drosophila*. *Annu Rev Genet.* 29:577-605.
- Wilson, R.K. y E.R. Mardis. 1997. Shotgun sequencing. In *Genome Analysis: A laboratory manual* (Editado por Birren B., Green E.D., Meyers R. M., Roskams J.). Cold Spring Harbor Press. Cold Spring Harbor, NY. 1:397-454.
- Wines, D.R. y S. Henikoff. 1992. Somatic instability of a *Drosophila* chromosome. *Genetics.* 131:683-91.
- Wong, L.H., K.H. Brettingham-Moore, L. Chan, J.M. Quach, M.A. Anderson, E.L. Northrop, R. Hannan, R. Saffery, M.L. Shaw, E. Williams y K.H.

- Choo. 2007. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res.* 17:1146-60.
- Zhang, P. y A.C. Spradling. 1995. The *Drosophila* salivary gland chromocenter contains highly polytenized subdomains of mitotic heterochromatin. *Genetics.* 139:659-70.
- Zhou, M., A. Bhasin y W.S. Reznikoff. 1998. Molecular genetic analysis of transposase-end DNA sequence recognition: cooperativity of three adjacent base-pairs in specific interaction with a mutant Tn5 transposase. *J Mol Biol.* 276:913-25.
- Zhu, J., H. Wang, J.M. Bishop y E.H. Blackburn. 1999. Telomerase extends the lifespan of virus-transformed human cells without net telomere lengthening. *Proc Natl Acad Sci U S A.* 96:3723-8.

Publicaciones

8. PUBLICACIONES

Una parte de los datos obtenidos durante la realización de esta tesis doctoral ha sido incluida en las siguientes publicaciones:

Mendez-Lago M, Wild J, Whitehead SL, Tracey A, de Pablos B, Rogers J, Szybalski W y Villasante A. Novel sequencing strategy for repetitive DNA in a *Drosophila* BAC clonereveals that the centromeric region of the Y chromosome evolved from a telomere. *Genome Res.* Enviado.

Villasante A, de Pablos B, Méndez-Lago M y Abad JP. Telomere maintenance in *Drosophila*: rapid transposon evolution at chromosome ends. *Cell Cycle.* 2008. 7: 2134-8.

Villasante A, Méndez-Lago M, Abad JP y Montejo de Garcíni E. The birth of the centromere. *Cell Cycle.* 2007. 6: 2872-6.

Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature.* 2007. 450: 203-18.

Villasante A, Abad JP, Planelló R, Méndez-Lago M, Celniker SE y de Pablos B. *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res.* 2007. 17: 1909-18.

Andreyeva EN, Kolesnikova TD, Demakova OV, Mendez-Lago M, Pokholkova GV, Belyaeva ES, Rossi F, Dimitri P, Villasante A y Zhimulev IF. High-resolution analysis of *Drosophila* heterochromatin organization using SuUR Su(var)3-9 double mutants. *Proc Natl Acad Sci U S A.* 2007. 104: 12819-24.

Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F, Villasante A, Dimitri P, Karpen GH, y Celniker SE. Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. *Science.* 2007. 316: 1625-8.

Villasante A, Abad JP y Méndez-Lago M. Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc Natl Acad Sci U S A.* 2007. 104: 10542-7.

Anexos

9. ANEXOS

ANEXO I. Análisis de los retrotransposones teloméricos presentes en el BACR26J21, a partir de la penúltima repetición de la unidad del satellite 18HT.

Las unidades de repetición del satélite 18HT aparecen en fucsia y en rojo, y corresponden a las dos últimas repeticiones del satélite en la región de transición del mismo. Dado que la unidad de repetición está formado por fragmentos de cuatro elementos teloméricos, se les ha numerado del 1 al 4 (comenzando en la última repetición del satélite, en rojo), y se ha continuado la numeración de los distintos retroelementos teloméricos presentes, hasta 9. Estos números aparecen escritos en verde. La presencia de elementos transponibles- no teloméricos- y de la duplicación segmental 42A, se ha indicado, mediante su nombre, pero su secuencia ha sido eliminada. Las duplicaciones en los sitios de inserción de transposones ha sido marcada en naranja. La inversión que precede al elemento diver, ha sido subrayada. Hacia el final de la secuencia, ésta ha sido marcada en tres tonalidades de azul, para mostrar las unidades que fueron amplificadas. Finalmente, en las regiones 5' UTR de los elementos *HeT-A* completos se han marcado en verde los motivos característicos de 8-17 bp que derivan de la región 3' del elemento *HeT-A* anterior al *HeT-A* completo a partir del cual se retrotranscribe cada elemento *HeT-A*.

3'UTR *HeT-A* (18HT) (3101bp)

CATCCAGCGCCCGCAAGCCCC

AGCCAAAGTACAACAACACTACTTACCTGCAATGTCTCCAGAGGCTTCCAGTGACTCGGTGC
 TTCCGTCCTTCTGGCGGGGTACCTCAATTATAAATTGTAAACATACAATACAATATGATA
 ATGTTACCAGTCCTAATTACTGTCAAAAGCCTACGTTTACAATATACTAACTACTTTTAT
 ATTAATACTAGTATAATACTAATACTATGTCCAACCCCCAACTCACCACATGCAATGTT
 AAATTCTAAAATGCCAATTATTGTACCTATGTATTACAAACATTGTAATCAAAGGCAAAA
 TAAATTGTGGATGCGGAACAGAATTCATTCTGTCTCCGTACCTCCACCAGCAATGTTATA
 AA

3'UTR *HeT-A* (18HT)

GCTGATCCTGAAACGACACACAGAAAATCAACCAGGACATAGAAGATGAATCGGAAGA
 TCGATGTGAAAAGGATTAGCGCGGAAGAAACATGATGAATCTAAGGCGACTCACTGTAGC
 AATATATGCACACTGTCACTTAACTGAATTTTCTTGCCGCGCGTTCTTTTGAAATATCC

CTTCTCCGCTGCAATCTTCACAATGCTGCTGCCTCCACGTCGATCAGGACACTTTGCTGC
TCTCATAACTTCCAGCTTGACGAGCGCCAACAGCGAGTTGACGCTAAAACCTAAAAACA
AACAAATTAACAACAAATTAATACTAATAAATAAAATAAAATCAAACAAAACACTTACCT
CACTGACAGCAGCCAATTGCTGACCCACATTCAACGCAACAGACACAGGAGACAGGCCC
CGCAAGCGCAAAAACAAATTTGCCAATTTTTGCGATTTTAAATACAACAAATTGACAATTT
TAGAATGCCGTCTCCATCTCCTGATTCCACTGCCCTTATTAGGATCATTAGCGCGTCAGG
CTCAAAATCTATTTTTCTCGTCAGTACCGTATCTTCAACGAATTTCCGATAACCAGCAA
TGAAAGGATAATTAATAAGAATGCGATACAAAAATTAATTAATGACATATAGATAATAGC
ATACCGGACAGACAAAGAAGCAGATGCCAATAGGCGGCTCCATCCTGCTGACGACAAACA
CGCAACTCCTTCTTCCAAGACTACAATTGCTGAAACAAGGGAACACAAGCTAATACTGGG
AAATATAAATTTAAATATAATAATTATCTAATTGCCACCTCGATAATTCCAAAACCGCG
GCATCCAGCTGCGACGGCACAGAAATATAGGTCTCCAAATTAGTCTTACAAACGTTT
CAGAAAAATAAAATGAATGCAATCATAAAAATAAACATAAATACTTACCTCCAGATTAGC
TTCCACCTTACGTACCTGAAAAACAAAAACAAAAATTAATGCAATTATTAAATTAACATA
ATACAAATATAATACTTACCCTCAAATTACCTCCCGTAAATTAACCTAAAAAAACAAAA
ATTAATACAATCTTAAAAACAAATAAGAAATGTAATACTTACCAAATTTAAATTTGTAT
TCATTTCCATGGCCCCCTATTGTTACGTCCTCGGCAACAAATCCTGTTCCGGCGGCTCCAA
GCTGCCAATCGCAACGCATTGGCCACAAGACACGGCGCTCCCGGCAACTCTCGGTGAACA
ACCGAGCTGATTTATGCCAGCAGCTGCGCAGGATCCATCTTCAGAATTTCTCTTCTCTGA
CGACCGGCGAAGCTGCCCTGCAATTCAATAAATTATTAAACAATTGCAAAACATCGGCCAC
TGAGGATAGAAGAGATATTCACCAATGACCGCGGGCGGGATGCTACCACCACGAATAA
ACTACCTGCAACACCGGCCGACATACATGTTGCAAGTGGCGCACTTCCAGGGGCCGCAA
CAAACCACAGCCAGCAAACAACAACAACTTACCTGTGATTACTCTCTGAGGCTTCCAGCGA
CTCGGTGCTTCCGTCTTTCTGGCGGGGTACCTGAAAAGAATTAAATCAAACAATGTTAG
TCTTAAATTTCAATAAATAAAAAAATACTTACCCTTAAATTAATACTAAATCTGT
GTCCAATCCCCAACTCACCCACGCAATGTACACCTCAAAATTCAAATAATTGTACCT
ACATATTGCACACACTGTAATCAAAGCAAAATATATTGTGGGTGCGGAACAGAATCATT
CTGTCTCCGTACCTCCACCAGCAAAGTTAAAAAA

3'UTR *HeT-A* (18HT)

GCAGAAATTAAGTGTGGATGCGGAA
TCATTCTGTCTCCATACCTCTACCAGCAAAGTTAAAAAA

3'UTR *TART-A* (18HT)

GACTGCGGCAGATGTCTTCCA
TCCCTCCTCAAAAGGTTGTGCGTTCACCAAAGTTGACGGGCCACGCCGGTGCATGCTTGC
TGCTACTGGCATCTCCGCAACAACGACGTTACCAGCACCCTGAAAGCAGAAACAGGTT
TACTATAAGCAATGCGCTGAACAACAAAACAAAGCAGAAACAGGTTTACTATAAGCAACG
CGCAGAAAAACAAATATTACAATAGGAAATAGCACGCGTTTTTCTGCGTCGTTTCCGCT
TAAGAGAAGAGAGGCCATTTTCCCTAAATGCCATTGACACTGGCAAACGCGAGTAACTGG
GAGAAATACTGGAATAAATTTAAATTTGAACAATTCTAGTATTTCTGGCCGAGAGGTAGC
TGGCGGTAAAAAAGATAACAGCCGATAAGGCCGCCAAAAAAGGCGATCATCCGCCGTGCC
TTTGAGCTATTGAAAAAGCCCGCCTACGCGCCTAGCCCAAATTTGCAAGTCAGTCGAGAT
TATAAAACCAACAGAGGCGGGCGCGCAACAGTAGCTAAAAGTGAAGCCAGCAGCAACAAC
AATAACTAATGGACGGACAGAACGTGAACCAAAGCGGAAGATGGGAATCGGTTTTATCCA
TCTCATCTGACTATGGTAACTGCTCATCCCCGCCGCCATCAACTATAGTCTTATCGCTGG
ATACCAAGTCATCGTTTAACTCTACTCTAAATTAATTACCAAACTATTTTCACCTATT
CCATAAGCTAATTCCTAAGTAATTTAATATTTAATTTTAAATTTCCACACTCGAACCATGG
CGCGCAAACTGCCGCCAAG

3'UTR *HeT-A* 1 (18HT) (2975 bp)

CATCCAGCGCCCCGAAGCCCCAGCCAAAGTACAACAATA
CTTATATACAATGTCTCCAGAGCCTTCCAGTGACTCGGTGCTTCCGTCTTCTGGCGGGG

TACCTCAATTATAAATTGTAAACATACAATACAATATGATAATGTTACCAGTCCTAATTA
CTGTCAAAAGCCTACGTTTACAATATACTAACTACTTTTATATTAATACTAATATAATAC
TAATACTATGTCCAACCCCCAACTCACCACATGCAATGTTAAATTCTAAAATGCCAATT
ATTGTACCTATGTATTACAAACATTGTAATCAAAGGCAAAATAAATTGTGGATGCGGAAC
AGAATTCAATTCTGTCTCCGTACATCCACCAGTAAAGTTATAAA

3' UTR *HeT-A* 2 (18HT)

GCTGATCCTGAAACGAC
ACACAGAAAAATCAACCAGGACATAGAAGATGAATCGGAAGATCGATGTGAAAAGGATTAG
CGCGGAAGAAACATGATGAATCTAAGGTGACTCACTGTAGCAATATATGCACACTGTCAC
TTAACTGAATTTTCTTGCGCGCGTTCCTTTTGAAATATCCCTTCTCCGCTGCAATCTTC
ACAAATGCTGCTGCCTCCACGTCGATCAGGACACTTTGCTGCTCTCATAACTTCCAGCTT
GACGAGCGCCAACAGCGAGTTGACGCTAAAACCTAAAAACAACAATTAACAACAATTA
AATACTAATAAATAAATAAATAAATAAACAACAACACTTACCTCACTGACAGCAGCCAATTG
CTGACCCACATTCAACGCAACAGACAACAGGAGACAGGCCCGCAAGCGCAAAAACAAAAT
TGCCAATTTTGTGCGATTTTAAATACAACAATTTGACAATTTTAGAATGCCGTCTCCATCT
CCTGATTCCTGTCCTTATTAGGATCATTAGCGCGGAGCTGACGCTCAAAATCTATTTT
TCTCGTCAGTACCGTATCTTCAACGAATTTTCCGATAACCAGCAATGAAAGGATAATTA
TAAGAATGCGATACAAAAATTAATTAATGACATATAGATAATAGCATACCGGACAGACAA
AGTAGCAGATGCCAATAGGCGGCTCCATCTGCTGACGACAAACACGCAACTCCTTCTTC
CAAGACTACAATTGCTGAAACAAGGGAACACAAGCTAATACGGGAAATATAAATTTAAA
ATATAACTTATCTAATTGCCACCTCGATAATTCAAAACCGCGGCATCCAGCTGCGA
CGGCACAGAAATATAGGTCTCCAAATTAGCTCTTACAAACGTTCCAGAAAAATAAAAAT
GAATGCAATCATAAAAATAAACATAATACTTACCTCCAGATTAGCTTCCACCTTACGTAC
CTGAAAAACAAAAACAAAAATTAATGCAATTATTAATTAATACTAAATACAAATATAATAC
TTACCTTCAAATTACCTCCCAGTAAATTAACCTCAAAAAACAAAAATTAATACAATCTTA
AAAACAATAAGAAATGTAATACTTACCAAATTTAAATTTTGTATTCAATTCCATGGCCC
CTATTGTTACGTCTCGGCAACAAATCCTGTTCCGGCGGCTCCAAGCTGCCAATCCCAAC
GCATTGGCCACAAGACACGGCGCTCCCGGTACTCTCGGTGAACAACCGAGCTGAAATTT
ATGCCAGCAGCTGCGCAGGATCCATCTTCAGAATTTCTCTCTCCTGACGACCGGCGAAGC
TGCCCTGCAATTCAATAAATTATTAACAATTGCAAACATCTACCACTGAGGGTAGAAGA
GATATTCACCCAATGACCGCGGCGCGGGATGCTACCACCACGAATAAATACTGCAACA
CCGGCCGGACGTACATGTTGCAAGTGGCGCACTTCCAGGGGCGCAACAAACCACAGCCA
GCAACAACAACAACACTTACCTGTGATTACTCTGAGCCTTCCAGCGACTCGGTGCTTCCG
TCTTTCTGGCGGGGTACCTGAAAAGAATTAATCAAACAATGTTAGTCTTAAATTTCAA
TAAATAAAAAAATACCTACCCTTAAATTAATACTAAATCTGTGTCCAATCCCCAAC
TCACCCACGCAATGACACTCAAAATTCAAATAATTGTACCTACATATTGCACACAC
TGTAATCAAAGGCAAAATATATTGTGGGTGCGGAACAGAATCATCTGTCTCCGTACCTC
CACCAGCAAAGTTAAAAAA

3' UTR *HeT-A* 3 (18HT)

GGCAAAATTAAAGTGTGGATGCGGAATCATCTGTCTCCGT
ACCTCTACCAGCAAAGTTAAAAAA

3' UTR *TART-A* 4 (18HT)

GACTGCGGCAGATGTCTTCCATCCCTACTCAAAGG
TTGTGCGTTTACCAAAGTTGACGGGCCACGCCGGTAGATGCTTGCTGCTACTGGCATCT
CCGCAACAACGACGTTACCGGCACCACTGAAAGCAGAAACAGGTTTACTATAAGCAATGC
GCAGAACAAACAAACAAAGCAGAAACAGGTTTACTATAAGCAATGCGCAGAAAAACAAAT
ATTACAATAGGAAATAGCACGCGTTTTTCTGCTTCGTTCCGCAAAAGAGAAGAGAC
CATTTTCCCTAAATGCAATTGACACTGGCGTTATGCGAGTAACACGGGAGAAATAGCTGG
AATCCCCAAAAATACTAATTTAAACAATTCTAGTATTTCTGGCCGAGAGGTAGCTGGCGGT
AAAAAGATAACAGCCGATAAGGCCGCCAAAAACGAGATCATCCGCCGTGCTTTGAGC

TATCGAAAAAGGAGGATGGGAATCGGTTTTATCCATCTCATCTGACGATGGTAACTGCTC
 ATCCTCTCCGCCATCAGCTATAGTCTTATCGCTGGATACCAAGTCATCGTTAACTCTAT
 TCTAAAATTACTTACCAAAACTATTTTACCTATTCCATAAGCTAATTCCTAAGTAATGT
 AATATTTAAATTTAAATTCACACTCGAACCATGGCGCGCAAAACGCGCCCAAGTATTT
 CAAAAAGGTCTCCCGCAAATTTAAATTATAAATTAGTACTTGTAACTTTGGCGCGCATAT
 CTGCGCAAATGTCACTATATATTTTTTAAGCAAACAAAAAAAAAAAA

3' UTR *HeT-A* 5

CCAAATAACAAA
 TATAATACTTACCAAATTTTAAATTTTGTGCGACTGTCTCGGCAACAAATCCTGTTCCG
 GCGGCTCCAATGCTGCCAATCCCGGCGCACTGGCCACAAGACTACAATTTCCATGACGAC
 TCCTCTGTGTCATGATGAGACAGATGACACCACCATAATGTCAGCAGCTCCAAAACAATACA
 ACGACGGCTCCGCGGAACCCATCTTCAGAATTCCTCTTCTGACGACCGGCGAAGTTGC
 CCTGGAATTCAACAATTTATTAATAATTGCAAACATCTACCACTGAGGGTAGAAGAGAT
 ACTCACCAAATGACTGCGGCGTGGGATGCTACCACCACGAACAACTAGCTGCAACGCCG
 GCCGGACATACGTGTTGCAAGTAGCGCGCATCCAGCGCCCGCAACATAGCCCCAGCCTAA
 GTACAAAAACTACTTACCTGCAATGTCTCCAGAGGCTTCCAGCGACTCGGTGCTTCCGTC
 TTTCTGGCGAGGGTAGCTGAAAAGAAACAAATTAACAATGTTAGACCTAAATTTCAATC
 TTTTGTAAAATAATTCAATTTGTTAAATGTAAACAAACCTGGCAATACGTTAATGTTAC
 CAGTCCATGTTACTGTCTAAAGCTAAGTTTACAAAATATACTAATTATAAACTAACTCG
 TGGTGAGTGGTGCCCAAGCCCCAACTCACCCCATGCAATATAACTCTGACTCCGTAC
 CTCCACCAGCAATGTTAAA

HeT-A 6

CTCCACCAGCAAAGTTA TTCAGCAAAGTATAA TAGTTAAAA C
 CAGCAAAGTTAAA AAAAAATTAAAAATAAAAAATTAAACAAATTAACAAAAATAAAAT
 AATTAATTTAAATTTATTTAACAAAATAATCCGCGCTTTCACCTGCAAAACATTCTGACG
 CGCAAATCTAATTTAAATCGCCTCTTCGTTTTGACAAATTAAGTTTAAATTTGTCTTC
 CGGCGCACAGTTAAAACCGCGACAACAATAAACATTTAGTGGACAAACGAAGGGCGAAC
 AATTAATTTGTGTAATATTTGTGCAAATTGACAAGGTGCCGCCATAACCAAAATGAGAAG
 AAAATCCATAGACGAAAAGAAGAAAGAAGTCCGCAACCAGGAAAAATTAACAAAAAAGGA
 AAGGAGGAAAACTAATAAAGGAGACGAACATAAATAGAGGAGACACCACAAATTAAGG
 CCAGGGTATTTATACCAACAAGTATCCCAAAAACGACTTTTTGGTTCGATTTTCAACCCCA
 AGCCGG

mdg1 (1-7451)

HeT-A (1488)

CCGGACATTTCTAAGCTTCGTTTAAATAAGAAACCCACCAACCGCTCTGGA
 AATACTGGGAAAAGAAGCATTTCCCTCACCAAAGGAGTGCTTCTTACGCTCTTCTGCT
 CAGGTTGATTCAAATTTAAATCCCAACATTAGCGCCATGCCCACTGTGGGGAATTTACCA
 GCAGCCCCGACCCCTTAGCCGACTGCTGCAAAGCGGGATTTATTTAATTCATCCTCCAAG
 AGCCCAACGAGCATCATATGAGTTTTTCGGAAGTGGTAGCTGGAGCAAGTCCAGTTTCT
 GTGGCACCCCTCTAATCCGGCACCCAGCAGAAAAGTCCAGGCAAGCGGACAAATGACGAT
 CTGGACTGCTCAAAATTTAAGACGCCCAATAAAAAAGTATGCGCGACATCCAACCCCTA
 CCCCCTCAAAAATACAGATGCAGAGCTGCCTCCATGGAAAACCGTGCCCCAGAGACGTAGA
 GCACCCACTATACTCGTCGATGATGTGAAGGAAATGTCCCACTACTGGAAGAGCTGAAC
 TGCACAGCAGGAGTCTCCAGCTATACCACCAGGGATATAGAAGTAAACGGGGTCAGGATC
 CATGCCAAGTA (2266)

TCTCTTCTACCC

HeT-A

(5531) TTCCAGGGCAACTTCGCCGGTCGTCAGGAAGAGGGAA
TTCTGAAGATGGGTTCGCGCAGCCGTCGTTGTATTGTTTTGGAGCTGCTGACATTATGG
TGGTGTCTATCTGTCTCATCATGACAGAGGAGTCGTCATGGAAATTGTAGGATTGGATTTG
CAGCG (5373)

CAACGCATGATGTCCAGTTCAGTACGGAGGCAGTTTAGGTGGCGCTCTTATGTA
GTGGAGCTGTGTGAAGGGGT

HeT-A

(3189) GCATGTCGTGGTATTTTTCCAGATGTCTTTGCGTTGGGT
TGCTTGCTTTGGTAAAGGCCGTGGCTTTGCCTTGTCTTTGAAGCTATCCTACCCTTTTT
GGTATTTTATTGGTGTTTTGGCGCCATTTATTAGTCGCTTCTGAATGGCCGTGTCATG
GGAGGAGAAACGTGGT (3014)

diver (1-6026)

HeT-A

(2988) TTCCGGCAAAATATGAACGACAGAAGACCACGTTTCTCTCTCCCATG
ACACGGTCATTTCAGAAGCGACCGAATAAATGGCGCCGAAACACCAATAAAATACCCAATA
AGGGTAGAATAGCTTCAAAGAACAAGGCAATGCCACGGCATTACCAAGACTAGCAACCC
AGCGCAAAGAAATCTGGAAAAATACCAGGACATGCAACCCCTTCACATAGCTCCACTACA
TAAGACGCGCCACCTAAGCTGCCTCCGTACT

F (821-4706)

HeT-A

CGAGCGCCAACAGCGAGTCTGCGC
TAAA

(4867)
AACTAAAAAGCTAACTCTGGGAAATATATATCTAGAATATAATACTTATCTAATTG
CCAATAAATGATTCCAAAAGCGCTGCATCCCAGCTGCAATGGCACAAAAATGTAGGGCC
TCCAATCAGCTCTCACAAAACGTGCCTGAAAAATAAAAACTAATGCAATCACAAAAA

(5962)
TAA
TCATAAATTACTTCAAATAAATTCAAATAATTCCACCTATATATTGCACACACTGTAATC
AAAGGCAAAAAATAAATCGTGGATGCGGAACAGAATTCACTCTAACTCCGTACCTCCACCAG
CAAAGTTA

HeT-A 7

TCCAGCAAAGTTTAA TAGTTGAA TCCAGCAAAGTTTAA AAACGTTAAA AATA
AAAAATTAAACAAATTAAAAAAAATAAAATAAATTAAACAAATAATTAAATTAAATTTAT

TTAACAAAATAATCCGCGCTTTCACCTGCAAAACATTCTGACGCGCAAATCTAATTTAAA
 TCGCCTCTTCGTTTTGACAAATTAAAAGTTTAAAATTGTCTTCCGGCCGCAAAGTTAAAA
 CCGCGACAACAATAAACATTTAGTCGACAAACGAAAGGCGAACAATTAATTTGTGTAATA
 TTTGTGCAATTTGACAAGCTGCCGCCATAACCAAAAGGAGAAGAAATTCATAGACGAAAT
 GAAGAAAGAAAACCGCAACCAGGAAAATTAAGGAAAGGAGGAAAACTAATAA
 AGGAGTTTTTCTCTCGAAAACAGCGATAGCATGGACAAAACCTCAGCTCGTCCACCAAAG
 TTGGACCCAATGCTCTTTCCCTCATTTTAATGCACACACATAAACAAAATAAGCCAACA
 ATAAAAATTTAGAAAGCCGCTCTAAATTTCCCGCGCTTGACAATTCTGACGCACATTACA
 AAAAAGCCAATGCGAGTAACAGTGGGTGGATATTGAAAGGAAAAACAGGACACGGTCCAG
 TGCTACAGATGCCAGGGCTTCCGGCATGTTAAAAATTCATGCATGAGGCCGCCAAGATGG
 ATGAAATGCGCTGGCGACCACCAATCATCTGTGCGCAAAACCAAGAACACCCCGCC
 ACCTGCGTAAACTGCTTTGGAGGGCACATTAGTGCTTACAAGGGATGTCCCGCTTACAAG
 GCCGAAAAACAAAAGCTTGCGGCAACAACATTGACATAAACAAAATACAGACAATTATT
 AAAGGCGCAACACATAACCTTCATAGACGTCAAGGCCCTCCTCCACGTAATAACACCCCT
 CGGATGCCGCACAGCTCAGCAATCCTGAGCAGATCAATCGCCGAAGCTCGCCAAGAGGCA
 GCCAGAAAGTCGATGTTAAATCCTTTCCGGCAAAATATGAACGACAGAAGACCACGTTTC
 TCCTCCCATTGACACGGCATTTCAGAAGCGACTGAATAAATGGCGCCGAAACACCAATAAA
 ATACCCAAAAAGGGTAGAATAGCTTCACAGAACAAGGCAATGTCACGCGATTTACCAAGA
 CTAGCAACCCAGCGCAAAAGAAATCTGAAAAATACCAGGACATGCAACCCCTTCACATAG
 CTCCACTACATAAGGCGCGCCACCTAAGCTGCCTCCGTACTGAACTGGACATCATGCGTT
 GCGCTGCAATCCAATCCTATTGACGAGCGCC

F (616-4706)

HeT-A

CGAGCGCCAACAGCGAGTCTGCGCTAAAAA

CTAAAA

(4867)

AGCTAACTCTGGGAAATATATATCTAGAATATAATACTTATCTAATTGCCAACT
 AAATGATTCCAAAAGCGCTGCATCCCAGCTGCAATGGCACAAAAATGTAGGGCCTCCAAT
 CAGCTCTCACAAAACGTGCCTGAAAAATAAAACTAATGCAATCACAAAAA

(5962)

TAATCATAA

ATTACTTCAAATAAATTCAAATAATTCCACCTATATATTGCACACACTGTAATCAAAGGC
 AAAATAAATCGTGGATGCGGAACAGAATTCACTCTAACTCCGTACCTCCACCAGCAAAGT
 TA

HeT-A 7

TC
 CAGCAAAGTTTAA TAGTTGAA TCCAGCAAAGTTTAA AAACGTTAA AAATAAAAAATT
 TAAACAAATTAAGGAAATTAATAAATTAACAAATAATTAATTAATTTATTTAACA
 AAATAATCCGCGCTTTCACCTGCAAAACATTCTGACGCGCAAATCTAATTTAAATCGCCT
 CTTCTGTTTGGACAAATTAAGGTTTAAATTTGTCTTCCGGCCGCAAAGTTAAACCGCGA
 CAACAATAAACATTTAGTGGACAAACGAAGGGCGAACAATTAATTTGTGTAATATTTGTG
 CAATTTGACCTGCTGCCGCCATAACCAAAAGGAGAAGAAATTCATAGACGAAATGAAGAA
 AGAAAACCGCAACCAGGAAAATTAAGGAAAGGAGGAAAACTAATAAAGGAGAC
 GAACTAACTAGAGGAGACACCACAAATTAAGCCAGGGTATTTATACCAACAAGTATC
 CCAAAAACGACTCACACTAATGCAGCTGCACTCACTCACACTGACGTACTCGAAGAGAG
 AAAATACCCACCTCAGCGATCTGCACTAATGCAGCTGCAGCCATTAAGGAGACACAGAT
 TTAATAGGCGTCAAATGAGCAACGGTATGGAATTTTCCCTCTCTCGAAAACAGCGAT
 AGCATGGACAAAACCTCAGCTCGTCCACCAAAGTTGGACCCAATGCTCTTTCCCTCATT
 TTAATGCACACACACAAACAAATAAAGCCAACAATAAAAAATTTAGAAAGCCGCTCTAAAT

TTCCCGCGCTTGACAATTCTGACGCACATTACAAAAAGCCAATGCGAGTAACAGTGGGT
 GGATATTGAAAGGAAAAACAGGACACGGTCCAGTGCTACAGATGCCAGGGCTTCCGGCAT
 GTTAAAAATTCATGCATGAGGCCGCCAAGATGGATGAAATGCGCTGGCGACCACCAATCA
 TCCTGTTGCGCAAAACCAAGAACCACCCCGCCACCTGCGTAAACTGCTTTGGAGGGCAC
 ATTAGTGCTTACAAGGGATGTCCCGCTTACAAGGCCGAAAAACAAAAGCTTGCGGGCAAAAC
 AACATTGACATAAAACAAAATACAGACAATTATTAAAGGCGCAACACATAACCTTCATAGA
 CGTCAAGGCCCTCCTCCACGTAATAACACCCCTCGGATGCCGCACAGCTCAGCAATCCTG
 AGCAGATCAATCGCCGAAGCTCGCCAAGAGGCAGCCAGAAAGTCGATGTTAAATCCTTTG
 CGGCAAAATATGAACGACACAAGACCAGTTTTCTCCTCCCATGACACGGCCATTTCAGAAG
 CGACTGAATAAATGGCGCCGAAACACCAATAAAATACCCAAAAAGGGTAGAATAGCTTCA
 CAGAACAAGGCAATGTACGGCATTACCAAGACTAGCAACCCAGCGCAAAAGAAATCTGG
 AAAAAATACCAGGACATGCAACCCCTTCACATAGCTCCACTACATAAGACGCGCCACCTAA
 GCTGCCCTCCGTACTGAACTGGACATCATGCGTTGCGCTGCAAATCCAATCCTATTGACGA
 CGCC

F (616-4706)

HeT-A

GAGCGCCAACAGCGAGTCTGCGCTAAAACTAAAA

(4867)

AGCTAACTCTGGGAAATATATATCT
 AGAATATAATACTTATCTAATTGCCAACTAAATGATTCCAAAAGCGCTGCATCCCAGCTG
 CAATGGCACAAAAATGTAGGGCCTCCAATCAGCTCTCACAAAACGTGCCTGAAAAATAAA
 AACTAATGCAATCACAAAA

(5962)

TAATCATAAATTACTTCAAATAAATTCAAATAATTCCACC
 TATATATTGCACACATTGTA

(6008)

-1731 (4599-1)

HeT-A

(6002) CATTGTAATCAAAGGCAAAATAAATCGTGATGCGG
 AACAGAATTCACTCTAACTCCGTACCTCCACCAGCAAAGTTA

HeT-A 7

TCCAGCAAAGTTAA TAGTTG
 AA TCCAGCAAAGTTAA AAACGTTAA AAATAAAAATTTAAACAAATTA
 AAATAAATTAACAAATAATTAATTAATTTATTTAACAAAATAATCCGCGCTTTCACC
 TGCAAAACATTCTGACGCGCAAATCTAATTTAAATCGCCTCTTCGTTTGCATAATTA
 AGTTTAAATTTGTCTTCCGGCCGCAAAGTTAAACCGCGACAACAATAAACATTTAGTGG
 ACAACGAAGGGCGAACAATTAATTTGTGAAATATTTGTGCAAAATGACAAGCTGCCGCC
 ATAACCAAAAAAGAGAAGAAAATACATAGACGAAATGAAGAAAGAGACCGCAACCAGGAA
 AATTTCAAAAAAAGGAAAGGAGGAAGAACTAATAAAGGAGACAACTAACTAGAGGAGA
 CACCACAAATTTAAAGCCAGGGTATTTATACCAACAATTATCAAAATTTTATACATTTAA
 AATCAGCAAAACCATGTCCAACAACCTACTTACCTACAACCTACATTTATCGCCCATGT
 CCCACGGATCAGAAAATTTCTCTGCTTAACACTATCATTATTAATCAGAAGAAATGCCCC
 CAAAGCCAAATAAAAGTTCAAAAACCTCTTCTGGGGCTGCCATAAAAAATGTTAATTCCT

TTCACATAAGGGGAAAAGAGAACACAAACACCAATAATGCCCAAAAAGACCCCTCTGGCTC
 ACCAATACTGTTGCAGTCTTTGGTGGTGCCAAAAGCAGCATCTCGAAGGGGAATTCGCCT
 TCTCTCCATTCTCCTCACATACACGCAAAGGGAAATTAGTAGCAGATTTAATAGGCGTC
 AAACGTAGCAACGGTATGGAAATTTTCCCTCTCTCGAAAACAGCGCTAGCATGGACAAA
 ACCTCAGCACCCCTTAGCCGGCTGCTGTAAAGCGGGATTTATTTAATTCATCTCTCCAAGA
 GCCCCAACTAGCATCATATGAGTTTTTTCGGAAGTGGTAGCTGGAGCAGGTCCAGTTTCTG
 TGGCACCCCTCTAATCCGGCACCACCGACGAAAACCTCAGGCAAGCGGACAAATGACGATC
 TGGATTGCTCAAATTTTAAAGACGCCCAATAAAAAAGTATGCGCGTCATCCAATTTTGTA
 CTTCCAGCATTTTTCTCCGACATCATACTCGTTTTTCAAAAAGCAAGGCAGCTCAATCTG
 TTTATGAAATGGACCCCCCGCCAAAGCAAGAAATGGACCCCTCGCCAGCCGTACCCCTG
 CAGCATCAATGTCTCTGCTCGCAGCGCACCCGGCGCCACCCGGGATCGCTCCCTACCCCC
 TCAAAACACAGATACAGAGCTGCCTCCATGGAAAATTATGCCCCAGAGCCGTAGAGCGCC
 CCTTACTCTCGTCAATGATGTGAAGGAAATTGTCCACTACTGGAAAAGCTGAACTACAC
 AGCAGGAGTCTCCAGCTATACCACAGGGCTATAGAAGTAAACGGGGTCAGGATCCAGGC
 CAAGGATATGACCGCTACAACAAAATTAAAGAAGTCTGGTGGCCAACGTATTTCTTTT
 ATTCACAAAACAAGCCCAAGTCCGAGAGAGGCTTCCGCGTCATCATACACCTCCATCA
 TTCCACACCATGCTCGTGGATAGTCGAGGAGCTGCTGAAGCTCGGATTTCCAAGCGCGCTT
 CGTCAGAAAACATGAAGAATCAGCTACAGGTGGTCCCATGCGAATGTTTGAAGTGGAGAT
 CATCATGCCCCAAGGAAGGCAGCCATGAAAAAATTATCTCACTCAAACAAATCGGTGGGCA
 ATAGGTGGATATTGAAAGGAAAAACAGGACACGGTCCAGTGCTACAGATGCCAGGGCTTC
 AGGCATGTTAAAAATTATGTCATGAGGCCGCCAAGATGCATGAAATGCGCTGGCGACCAC
 CAATCACCCCTGTTGCGCCAAACCAAGAACCACCCCGCCACCTGCGTAAACTGCTCTGGA
 GGGCACATTAGTGGTTACTTTATTCAGTCGCTTCTGAATGGCCGTGTCATGGGAGGAGAA
 ACGTGGTCTTCTGTCTGTTTATTTTGGCGGAAAGGATTTAACATCGACTTTCTGTCTGC
 CTCCTGGCGAGCTTCGGCGATTGATCTGCTCAGGATTTCTGAGCTGTGTGGCATCCGAGG
 GGTGTTATTACGTGGAGGGGGCCCTTAACGCTCTATGAAGTTATGTGTTGCGCTCTTAAT
 TGTCCGATTTTTGTTTTATGTCAATGTTGTTTGGCGTAAGCTATTGTTTTTTCGGCCTTGTA
 AGCGGGACATCCCGCTATTGGCCGAAAAACAATAGCTTACGGCAAACAACCTGACTGACGT
 GCGGGATGCTAACACCACGAACAACTACCTGCAACGCCGGCCGGACATACATGTTGCAA
 GTAGCGCGCATCCAGCGCCCCCAACATAGCCCCAGCCTAAGTACAACAACCTACTTACCTG
 CAATGTCTCCAGAGGCTTCCAGCGACTCGGTGCTTCCGTCTTCTAGAGGGGGTACCTGA
 AAAGAAACAAATTAAACAATATTAGTTCTAAATTTCAATCTTTTTTGTAGAATAATTTAA
 ATTGTTAAATGTAAACAAACCTTGCAATACGTTAATGTTACCAGTCCATGTTACTGTCTA
 AAAGCTAAGTTTACAAAAATACTAATTATATACTAACCTACCACGCCCAAGCCCCAACT
 CAACCTTGCAATGTAATACTCATAAATTCAAATAATTGTACCTATATATTGCACACACT
 GTAATCAAAGGCAAAATAAATCGTGGATGCGGAACAGAAATTCACCTCTGACTCCGTACCTC
 CACCAGCAAAGTTGAA

HeT-A 6

CAAAGTTAAA AAATATTAGAAATAAAAAATTTAAACAAATTAAAAAAAATTAATAATT
 AAATTAAATTTATTTAAACAAATAATCCGCGCTTTCACCTGCAAAACATTCTGACGCGCA
 AATCTAATTTAAATCGCCTCTTCGTTTTGACAAATTAAAAGTTAAAAATTTCTTCCGGC
 CGCAAAGTTAAAACCGCGACAACAATAAACATTTAGTGGACAAACGAAGGGCGAACAATT
 AATTTGTGTAATATTTGTGCAATTTGACAAGCTGCCGCCATAACCAAAAGGAGAAGAAAT
 TCATAGACGAAATGAAGAAAGAAAACCGCAACCAGGAAAATTTAAAAAAGGAAAGGAGG
 AAAAACTAATAAAGGAGACGAACTAACTAGAGGAGACACCACAAATTTAAAGCCAGGGT
 ATTTATAACAACAAGAATCCCAAAAACGACTCACACTAATGCAGCTGCACTCACTCACAC
 TGACGCTACTCGAAGAGAGATAATACCCACCTCAGCGATCTGCACTAATGCAGCTGCAGC
 CATTAATACGGACACAGATTTAATAGGCGTCAAACCTAGAGCAACGGTATGGAAATTTCCC
 CTCTCTCGAAAAACAGCGATAGCATGGACAAAACCTCAGCTCGTCCACCAAAAATGGACCC
 AATGCTCTTTCCCTCATTTTAAATGCACACACATAAACAATAAAGCCAACAATAAAAAAT
 TTAGAAAGCCGCTCTAAATTTCCCGCGCTTGACAATTCTGACGCACATTACAAAAAGCC
 AATGCGAGTAACAGTGGGTGGATATTGAAAGGAAAAACAGGACACGGTCCAGTGCTACAG
 ATGCCAGGGCTTCCGGCATGTTAAAAATTCATGCATGAGGCCGCCAAGATGGATGAAATG

CGCTGGCGACCACCAATCATCTGTTGCGCAAAACCAAGAACCACCCCGCCACCTGCGT
 AAACGTCTTTGGAGGGCACATTAGTGCTTACAAGGGATGTCCCGCTTACAAGGCCGAAAA
 ACAAAGCTTGGCGCAAAACAACATTGACATAAAACAAAATACAGACAATTATTAAAGGCGC
 AACACATAACCTTCATAGACGTCAAGGCCCTCCTCCACGTAATAACACCCCTCGGATGCC
 GCACAGCTCAGCAATCCTGAGCAGATCAATCGCCGAAGCTCGCCAAGAGGCAGCCAGAAA
 GTCGATGTTAAATCCTTTCCGGCAAAATATGAACGACAGAAGACCACGTTTCTCTCCCA
 TGACACGGCCATTGAGAAGCGACTGAATAAATGGCGCCGAAACACCAATAAAATACCCAA
 AAAGGGTAGAATAGCTTCACAGAACAAGGCAATGCCACGGCATTACCAAGACTAGCAAC
 CCAGCGCAAAAGAAATCTGGAAAAATACCAGGACATGCAACCCCTTCACATAGCTCCACTA
 CATAAGACGCGCCACCTAAGCTGCCTCCGTACTGAACTGGACATCATGCGTTGCGCTGCA
 AATCCAATCTTATTGACGAGCGCC

F (616–4706)

HeT-A

CGAGCGCCAACAGCGAGTCTGCGCTAAA

(4867)

AACTAAAAAGCTAACTCTGGGAAATATATA
 TCTAGAATATAATACTTATCTAATTGCCAACTAAATGATTCAAAAAGCGCTGCATACCAG
 ATGCAATGGCACAAAAATGTAGGGCCTCCAATCAGCTCTCACAAAACGTGCCTGAAAAAT
 AAAAATAATGCAATCACAAAA

(5962)

TAATCATAAATTACTTCAAATAAATTCAAATAATTCT
 ACCTATATATTGCACACCCTGTAATCAAAGGCAAAATAAATCGTGGATGCGGAACAGAAT
 TCACTCTAACTCCGTACCTCCACCAGCAAAGTTA

HeT-A 7

TCCAGCAAAGTTTAA TAGTTGAA TCCAG
 CAAAGTTAAA AAACGTTAA AAATAAAAAATTTAAACAAATTAATAAAAAATATAAATAAA
 TTAAACAAATAATTAATTAATTTATTTAAACAAATAATCCGCGCTTTACCTGCAAAA
 CATCTGACGCGCAAACTAATTTAAATCGCCTCTTCGTTTTGACAAATTAATAAGTTTAA
 AATTGTCTTCCGGCCGCAAGTTAAAACCGCGACAACAATAAACATTTAGTGGACAAACG
 AAGGGCGAACAATTAATTTGTGAAATATTTGTGCAAAATTGACAAGCTGCCGCCATAACCA
 AAAGGAGAAGAAAATACATAGACGAAATTCGTCTATGAATACATAGACGAAAAAAGGA
 AAGGAGGAAGAACTAATAAAGGAGACAAACTAACTAGAGGAGACACCACAAATTTAAAG
 CCAGGGTATTTATACCAACAATTATCAAATTTTATACATTTAAATCAGCAAAACCCAT
 GTCCAACAACCTACTTACCTACAACCTACATTTATCGCCCATGTCCCACGGATCAGAAAA
 TTCTCTGCTTAACACTATCATTATTAATCAGAAGAAATTGCCCTCAAAGCCAAATAAAG
 TTCAAAAACCTCTTCTGGGGCTGCCATAAAAATTTGTTAATTCCTTTTACATAAGGGGAAA
 GAGAACACAAACACCAATAATGCCCAAAAATACCCCTCTGGCTCAACAATACTGTTGCA
 GTCTTTGCTGGTGCCAAAAGCAGCATCTCGAAGGGGAATTCGCCTTCTCCTCCATTCTCC
 TCACATACACGCAAAAGGGAATTTAGTCCCAAAAACGACTCACACTAATGCAGCTGCACTC
 ACTCACACTGACGCTACTCGAAGAGAGAAAAATACCCACCTCAGCGATCTGCACATAATGCA
 GCTGCAGCCATTAATACGGACACAGATTTAATAGGCGCAAAACTGAGCAACGCTATGGAA
 ATTTTCCCCTCTCTCGCAACAGCGATAGCATGGACAAAACCTCAGCTCATCCACCAAAA
 ATTTGACCCAATGCTCTTTCCCCTCATTTTAATGCACACAAACAAACAAATAAAGCTAAT
 AATAAAAAGTTAGAAAGCCGCTCTAAATTTCCCGCGCTTGACAATTCTGACGCACATTAC
 AAAAGCCAATGCTAGTAACAGTGGGGAAATCTTCTCCCACAAATACGAATTGACGAAC
 ATAAGCAAGAGGAAGGGCCATGTACTATTCTCAATGCTTTTTGGTCGATTTTCTAAGCGT
 AGTTTATTTTAGTTAAGCTTAGTTCTAAGCTTAGTTTAGTTTCCTTTATTCACAAACAAG
 CCCAAGTCCGAGAGAGGCTTCCGCGTCATCATTATACACCTCCATCATTCACACCATGC

TCGTGGATAGTCGAGGAGCTGCTGAAGCTCGGATTCCAAGCGCGCTTCGTCAGAAACATG
 AAGAATCCAGCTACAGGTGGCCCCATGCGAATGTTTGAAGTGGAGATCATCATGACCAAG
 GATGGCAGCCATGAAAAAATTATCTCACTCACAACAAATCGGTGGGCAATAGGTGGATAT
 TGAAAGGAAAAACAGGACACGGTCCAGTGCTACAGATGCCAGGGCTTCAGGCATGTTAAA
 AATTCATGCATGAGGCCGCCAAGATGCATAGAATGCGCTGGCGACCACCAATCATCTCTGT
 TGCACCAAACCAAGAACCACCCCGCCACCTGCGTAAACTGCTCTGGAGGGCACCATTAGT
 GCTTACAAGGGATGTCCCGGTCACAAGGCCGAAAAACAAAAGCTTGGCGCAAAACACATT
 GACATAAACAAAATACGGACAATTAAAGACGCAACACATAACCTTCATAGACGTCAAGGC
 CCCCCCTCCACGTAATAACACGTTGCGCTCCAAATCCAATCCTATTGACGAGCGCCAACAG
 CGAGTCTGCGCTAAAAACTGAAAAGCTAACTGTGGGGAATATATATCTAGAATATAATAC
 TTATCTAATTGCCAACTCAATGATTCCAAAAACGCTGCATCCAGCTGCAATGGCACAAA
 AATATAGGGCCTCCAATCAGCTCTCACAAAACGTGCCGAAAAATAAAAACTAATGCAAT
 GACAAAAATAAAAAATAATCATAAATAACTTTTACCTTAAAGTACCTGAAAAACAAAAAC
 AAAAAATTAAATACAATTATTAAATCAAATAAATACAAATATAATACTTACCTCCAAATTA
 CCTCCCACCTAATGTACCTGAAAAACAAAAATTAATACAATATCAAAAAACAAATAACAA
 ATGTAATACTTCCCAAAATTTAATGTTGTATTTCATTTCCATGTTTCCAATCGTTGCGACT
 GTCTTCGGCAACAAATCCTGTTCCGGCGGCTCCAATGCTGCCAATCCCGACGCACGTGCC
 ACAAGACGCGGCGCTACTGGCAACTCTCGGTGAACAATCGAGCTACAATTAATGACGACT
 CCTCTGTACAGATGAGACAGATGACACCACCATAATGCCAGCAGCTCCAAAAACAATACAA
 CGACGGCTGCGCGGAACCCATCTTCAGAATTCCCTCTTCCCTGACGACCGGCGAAGTTGCC
 CTGGAATTAACAATTTTATTAATTGCAAACATCTACCCTGAGGGTAGAAGAGATACTCAC
 CAAATGACTGACGCGCGGGATGCTAACACCACGAACAACTACCTGCAACGCCGGCTGGA
 CATACATGTTGCAAGTAGCACGCATCCAGCGCCCCAACATAGCCCCAGCCTAAGTACAA
 CAACTACTTACCTGCAATGTCTTCAGAGGCTTCCAGCGACTCGGTGCTTCCGTCTTCTA
 GAGGGGGTACCTGAAAAGAAACAAATTAACAATATAAGTCATACATTTCAAGGTTTTTT
 TTTAAAAATAATTTAAATTGTTAAATGTAAACAAACCTTGCAAAAGGGTAATGTTACCAGT
 CCATGTACTGTCTAAAACCTAAGTTTACAAAAAGTACTAATTTAAGCTAACTCACAC
 GCCCCAAGCCCCAACTCACCCCTTGCAATGTAATACTCATAAATTCAAATAATTATACCT
 ATATATTACACACACTGTAATTAAGGCAAAATAAATCGTGGATGCGGAGGAGAATTCAC
 TCTGACTCCGTAACCTCCACCAGCAAAGTTAAA

HeT-A 8

CTCCACCAGCAAAGTTA TCCAGTAAAAATTTAA
 TAGTTAA ()AACCGCGACAACAACAACATTTAGTGGACAAACGAAGGGCGAACAATT
 AATTTGTGTAATATTTGTGCAAATTGACAAGCTGCCGCCATAACCAAAAGAAGAAGAAAA
 TCCATAGACGAAAAAGAAGAAGAACCGCAACCAGGAAAGTTAAAAAAGGAAGGAG
 GAAAAATAACAAAGGAGACGAACATAAAGTAGAGGAGACACCACAAATTAAGAACGGG
 TATTTATACTAACAAAGTATCAAATTTTTAAACCTACCTTTTTATACCTTTTTTCCCTGCAG
 CATCAATGTCTCTGCTCGCAGCGCAGCGGCCACCCGGGATTGCCCCCTAACCCCTCA
 AAATACTGATGCAGAGCTGCCTCCATGGAACACCGTGCCCCGTAGACGTAGAGCACCCAC
 TATACTCGTCAATGATGTGAAGGAAATTGTCCCACTACTGGAAAAGCTGAACTACACAGC
 AGGAGTCTCCAGCTATACCACCAGGGCTATAGAAGTAAACGGGGTCAGGATCCATGCCAA
 GGATATGACCGCCTACAACAAAATTAAGAAGTCCTGGTGGCCAACGGATTTTCTTTATT
 CACAAACAAGCCCAAGTCCGAGAGAGGCTTCCGAGTCATCATACACCTCCATCATTC
 CACACCATGCTCGTGGATAGTCGAGGAGCTGCTGAAGCTCGGATTCCAAGCGCGCTTCGT
 TAGAAACATGAAGAATCCAGCTACAGGTGGTCCCATGCGAATGTTGAAGTGGAGATCAT
 CATGACCAAGGATGGCAGCCATGAAAAAATTATCTCACTCAAACAAATCGGTGGGCAATA
 GGTGGATATTGAAAGGAAAAACAGGACACGGTCCAGTGCTACAGATGCCAGGGCTTCAGG
 CATGTTAAAAATTCATGCATGAGGCCGCCAACATGCATGAAATGCGCTGACGAACACCTG
 TCATCTGTTGCACCAAGTCAAGAACCACCCCGCCACCTGCGTAAACTGCTCTGGAGAG
 CACAGTGCTTACAAGGGATGCCCGCTTACAAGGCCGAAAAACAAAAGCTTGGCGGTAATC
 AACATTGACATAAAACAAAATACGGACAATTAAAGACGCAACCCATAACCTTTTATAGACGT
 CAAGGCCCCCTCCACGTAATAACACCCCTCGGCTGCCACACAGCTCAGCAATCCTGAGC
 AGATCAATCGCCGAAGCTCGCCAAGAGGCAGCCAGAAAGTCGATGTTAAATCCTTTCCGG
 CAAAATATGAACGACAGAAGACCACGTTTCTCCTCCCATGACACGGCCATTGAGAAGCGA
 CTGAATAAATGGCGCCGAAACACCAATAAAATACCAAAAAGGGTAGGATAGCTTCAAAG

GACAAGGCAATGCCACGGCCTTTACCCAAGACAAGTAACCCAGCGCAAAGACATCTGGAA
 AAATAGCAGGACATGCTCCGAAACGAAAGGAGTGAATAAACTGACCATGAACCTGAAAAA
 GGTACTTCTAATCCCAGCCGAGTCGGCAACGACAGCCCTCCGACCACCACCAGAGCCGCC
 AGAGCTAACCTTAAGCCAAGAATTATCGACTAAGCTACACCATCGCCAAGAAATTCCAAT
 CCTTACCTACAGAAAAGCTCGGACGACCCCCACCATAAACCTAGCTAATAGAGTCGACAAT
 TTAGAAAAAGTAAATTGACATTTTCATGGCTTTAATCATACAAGGAAGCAATAACAATAAT
 CTTGACATCGATACATCCATCTAAATTTACATTACATTTATTTATATTTATATCTATTAT
 ATCTATATATATATCCGACATAAAAGCGCACGTATGTCTACTTCCGTAATGCTCTACTTA
 TCATCACCTTCCCTCGACGCATAGCCTAAACCACCTGTTGAAAAACAAATCAATTAGATGGA
 TGTCATAAAAAACGTAAATAAAAAATCCTCTCACCTCAAGCATCCGGATAAAAAAGGTAAG
 ACAACGCACTCCAGCAGATTCTGTATGAAGCTATGTGAAGAAAACACACCAGGACTCAA
 AAGTCAAAATCGGAGGATGGACATGAGAAGAATTAGTGCGGCAGAACATGATGAATAGAG
 GCGACTCGCTGCAGCAAAATATGCACAACGCCACTTACCTGAATCTTCAGCGGCACAGTC
 TTTTTTATGTACCATTATCTCCGCCGCAACCCCTTCACACAGCTGCACTACATAAGACGT
 GCCACCTAAGCTGCCTCCGTACTGAATTGGACATCATGCGTTGCGCTGCAAAATCCAATCC
 TATCGACGAGCGCCAACAGCGAGTCTGCGCTAAAAACCAAAAAACAAAAATAAAACAAT
 TAATCAACAGCAAATTGAATATAACACTCAACAATAACAATCACTTACCTCCTTAACCTG
 CATCCAATCGCTGACTCAAATCCAACACAACCGACAACAGGAGACGGGTCCCGCAAAACGC
 AAAACAAAAATCGCCAACATTTGAGATTATAAATACAAAAATCTACAAGTTTGTGTATGCCG
 TCTCCATCTCCTGATGCCACTGCCCTAATTAGGATTATTAGCGCGGAGCTGACGCCACAT
 TAATAAGCTGCAAAATTTGTTCTCAAAATGTGTATTTCTCCTAAGTATTGTATCTTCAAC
 GAATTCTCCGCCAACCTGTAACGAAAGGAAAATTAACAACAATATATACAAAATTAATCA
 ATGACACAGAAAATAGCAAACCTGACAGGCAAAGTATCTGATGCAAAATAGGGCGACTCCA
 TCCTGCCGACGACAAGCACGCAACTTCTTCCTCCAAGCTTACAATTACTGAAAATAAGGAA
 ACACAAGCCAACCTCTCGGAAATATGTATCTAGAATATAAATACTTATCTAATTGCCAACTC
 AATGATTCCGAAAACGCGGGATCCCAGCTGGAATGGCACAAAAATATAGGCCCTCCAATC
 AGCTCTCACAAAACGTGCCTGAAAAATAAAAACTAATGCAAA

sd 42A

TTATAAATATTTTTGTGTTTATTTCCATCTTACTATTG
 TAGTAAGAAATATGTATATGAGTCCCAATATATTGTACATGAAATAATTGATACGAATGA
 GACCTACAGTAAAGTAGAATAAGAAGCAATGGAGGAATAGCCACAAATTTCCAAAATTCA
 CAAAAATAATCATAATACTTACCTCCAAATTAGCTTCCACTTAATGTACCTGAAAAACAA
 AAACAAAAAATTAATAAAATTAATAAATACACGAACAAACTACGTGCAACGCCGCGCCGA
 ACATACATGTTGCCATCGCTGGCGGAGGTACCTGAAAAGAAACCAATTAAACAATGTTAG
 TTCTAAATTTCAATCTTTTTGTGTAATAATTCAAATTGTTAAATGTAAACAAACCTTGC
 AATACGTTAATGTTACCCAGTCCATGTACTGTCTAAAAGCTAAGTTTACAAAAATACTAA
 TTATATACTAATCACCACGCCCAAGCACCAACTCACCCTATGCAATATAAATACTTATA
 ACTTTAAATAATTGTACCTATATATTGCACACACTGTAATCAAAGGCAAAAAATAATCGTG
 GATGCGGAACAGAATTCACCTCTGTCCCGTACCTCCACCAGCAAAGTTAAAA

HeT-A 9

A ATAATAAAATAAAATAAATAAATACTTAAATTAAAGTAATAAAAAATAACGCGCTTCG
 TCTGCAAAATACTCTCATGCGTAAATTTAATTAAATCGACTTTCAGTTTATAAGTTAA
 AAGTTTAAATTTGTCTTCCGGCCGCAAAGTTTGAACCGCGACATTTAAACATTTAATG
 GACAAACTAAAAAGCGAAAAATTAATCAGTGAATTATTTGTGCAAAATTGACAAGCTGCCG
 CCATAACCAAAAGGAGAAAGAAAAAGAGAGAAGAAAGCAGAAAAAGTAAACCTAAAGA
 AGAAAAGGAGGAAAACACAATTAAAGAGGCGATCTAAACTGGAGGACACATTACAATTAA
 AAGCCAGGGTATTTATACCTTCACAAGTATCGTTTTAATATAAAAAAACCAACATGTCC
 GTGTCGGCAACCTTTTCTCTGACGATTTGATCAATCCCAGAACAGCGATCTTCCCCGTT
 CTACCTCAATATATCGCCACGTCCACGGATCAGACAATTCTCTGCTTAACACTATCAT
 TATTAATCAGAAGAAATTGCCCTCAAAGCCAAATAAAAGTTCAAAAACTCTTCTGGGGCT

GCCATAAAAAATTGTTAATTCCTTTTCACATAAGAAGAAAGAGAACACAAACACTAATAAT
 GCCCCAAAAGACCCCTCTCACTCACCAATTCTACTGCAAGCACTTGTGGCGCCAAAAGC
 AGAATTTTCGAAGGGGAAATTGCCTTCTTCTCCGTTCTCCTCACACACATATGAGGGGAAA
 TTACTCTCAAAACTGACTCACACTGATGCTGCTGCCACACTCACATCGACGCTAATCGA
 AGAGTGAAAAATATTCACCTCAGCGATCTGCACTATTGCAGCTGCAGCCACAAAATACGGAC
 ACAGACTTCAAAAAGCGCCAAAACGAGCGATGCTATGGGAAGTTTCCCTCTCTCGCACAG
 TGACAATAGCATAGAGAAAAATCTTCCAATGCTTCTTCCCTCCTTCTCATGCACACACA
 CACACTAGCAAAGCCACTGACATAAGCTTAGAAAGCCGCTCAAAATGTCCCGCGCTTGCC
 AATACTGACACACGCTCCAAAAAAGCCAATGTTAGTGACAGTGCGGGAATATTCTCCCCA
 CTTATACAAAATTGACGAACGCAAGCAAGAGGACAGGCCTTGCACAACATATCAACGCTTTT
 TGGTCGATTTTTAAACCCAAACCGGACGTTACTAACTAAGTCTAAATAGGAAACCCACT
 AATCCCACTAGAAATACTGGGAAAAAAGCATTCCAACCTTGTAACTTCCAGCATTTTTTC
 CTCCGCTCATCATACCCGTTTTTCAAAGCAAGGCAGCTCAATCTGTTTATGAAATGGACC
 CCCCCGTCAAAGCCAGAAATGGACCCCCCGCCAGCCGTTACCTGCAGCATCAATGTCTC
 TGCTCGCAGCGCAGCGGCCACCCGGGATCGCCCCCTACCCCTCAAAAACAGATAC
 AGAGCTGCCTCCATGGAAAATCGTGCCCCAGAGCCGTAGAGCGCCCCCTATACTCGTCAA
 TGATGTAAAGGAAGCGCGCTTCGTGAGAACATGACGTATCCAGCTACAGGTGGACCCAT
 GCGAATGTTTGAAGTGGAGATCATCATGACCAAGGATGGCAGCCATGAAAAAATTATCTC
 ACTCAAAACAAATCGGTGGGCAATAGGTGGATATTGAAAGGAAAAACAGGACACGGTCCAG
 TGCTACAGATGCCAGGGCTTCAGGCATGTTAATAATTTCATGCATGACGCCGCCAAGATGC
 ATGAAATGCGCTGACGAACACCTGTCTCCTGTTGCACCAAGCCAATAACCACCCCGCC
 ACCTGCGTCAACTGCTCTGGGGACCATTTTAGTGCATACAAAGGATGCCCGCTTACAAG
 GCGGAAAAACAAAAGCTGGCGGCAACAACATCGACGTAAGCAAAATAAGAACAATCAAA
 GCGCAACAAAATAACTTTTATAGACGTCAAGCCCCCACCCCCACGCAATAACACCCC
 TCGGCTGCCACACAGCTCAGCAATCCTGAGCAGATCAATCGCCGAAGCTCACCAAGAGGC
 AGCCAGAAAGACGATGTTAAATCCTTTCCGGCAAAATATGAACGACAGAAGACCACGTTT
 CTCAATGGGCACAAATATATAGGGCCTCCAATCAGCTCTCACAAAACGTGCCTGAAAAATA
 AAAACTAATGCAATCACAAAAATAAAAAATCATAAATTACTTCCACCTAAAGTACCTGA
 AAAACAAAAACAAAAAATTAATACAATTATTAATCAAATAAATACAAATATAATACTTA
 CCTCAAAATTACCTCCAGCTAATGTACCTGAAAAACAAAAATTAATACAATATCAAAA
 ACAATAACAAATGTAATACTTACCAAATTTTAAATGTTGTATTTCATTTCCATGTTTCCAA
 TCGTTGCGACTGTCCTCGGCAACAAATCGTGTTCCGGCGGCTCCAATGCTGTCAATCCCG
 ACGCACTGGCCACAAGACGCGGCGCTACTGGCAACTCTCGGTGAACAACCGAGCTACAAT
 TTCTATGACGATAAGAGATACTCACCAAATGACTCCGGCGCGGGATGCTACCACCACGAA
 CAACTACCTGCAACGCCGGCCGTACATACATGTTGCAAGTGTGCGCATCCAGCGCCCCG
 CAACATAGCCCCAGCCTAAGTACAACAACACTTACCTGCAATGTCTCCAGAGGCTTCCA
 GCGACTCGGTGCTTCCGTCTTCTGGCGAGGGTACCTGGAAAGAAACAAATTAACAATG
 TTAGACCTGAATTTCAATCTTTTGTAAATAGTTCAAATGTTAAATGTAAACAAACCT
 TGCAATACGTTAATGTTACCAGTCCATGTTACTGTCTAAAAGTTAAGTTTACAAAAAAC
 TAATTGTAACTAATCACCACGCCCCAAGCCCCAACTCGCCCCATGCAATGTAATACTC
 ATAAATTCAAAAAAGCCAATGTTAGTGACAGTGGGGAATATTCTCCCCACTTATACAAA
 TTGACGACCGCAAGCAAGAGGACAGGCCTTGCACAACATCAACGCTTTTTTGGTCGATTT
 TTAAACCCAAGCCGACGTTACAGAATTCACCTCTGTCCCGTACCTCCACCAGCAAAGTT
 AAAA

HeT-A 9

GCAAAAGTTAA ATAATAAAATAAAATAAATAAATACTTAAATTAAAGTAATAAAAAAT
 AAACGCGCTTCGTCTGCAAAATACTCTCATTTATAGACGTCAAGGCCCCCTCCACGTAA
 TAACACCCCTCGGTGCCACACAGCTCAGCAATCCTGAGCAGATCAATCGCCGAAGCTCG
 CCAAGAGGCAGCCAGAAAGTCGATGTTAAATCCTTTCCGGCAAAATATGAACGACAGAAG
 ACCACGTTTCTCCTCCCATGATACGGCCATTGAGAAGCGACTGAATAAATGGCGCCGAAA
 CACCAATAAAAAATACCCAAAAAGGGTAGGATAGCTTCAAAGGACAAGGCAATGCCACGGCC
 TTTACCCAAGACAAGTAACCCAGCGCAAAGACATCTGGAAAAATAGCAGGACATGCTCCG
 AAACGAAAGGAGTGAATAAACTGACCATGAACCTGAAAAAGGTACTTCTAATCCCAGCCG
 AGTCGGCAACGACAGCCCTCCGACCACCACCAGAGCCGCCAGAGCTAACCTTAAGCCAAG

AATTATCGACTAAGCTACACCATCGCCAAGAAATCCAATCCTTACCTACAGAAAAGCTC
 GGACGACCCACCATAAACCTAGCTAATAGAGTCGACAATTTAGAAAAAGTAAATTGACAT
 TTTAATGGCTTTAATCATACAAGGAAGCAATAACAATAATCTTGACATCGATACATCCAT
 CTAAATTTACATTACATTTATTTATATTTATATCTATTATATCTATATATATATCCGACA
 TAAAAGCGCACGTATGTCTACTTCCGTAATGCTCTACTTATCATCACCTTCCTCGACGCA
 TAGCCTAAACCACTGTTGAAAAACAAATCAATTAGATGGATGTCATAAAAAACGTAAATAA
 AAAATCCTCTCACCTCAAGCATCCGGATAAAAAAGGTAAGACAACGCACTCCAGCCGATT
 CCTGATGAAGCTATGTGAAGAAAACACACCAGGACTCAAAAGTCAAATCGGAGGATGGA
 CATGAGAAGAATTAGTGCGGCAGAAGCATGATGAATAGAGGCGACTCGCTGCAGCAAAAT
 ATGCACAACGCCACTTACCTGAATCTTCAGCGGCACAGTCTTTTTTATGTACCATTATCT
 CCGCCGCAACCCCTTCACACAGCTCCACTACATAAGACGTGCCACCTAAGCTGCCTCCGT
 ACTGAATTGGACATCATGCGTTGCGCTGCAAATCCAATCCTATCGACGAGCGCCAACAGC
 GAGTCTGCGCTAAAAACCGAAAAAACAAATAAAACAATTAATCAACAGCAAATTTGAATA
 TAACAATCAAACAATAACAATCACTTACCTCCTTAACGTCATCCAATCGCTGACTCAAAT
 CCAACACAACCGACAACAGGAGACGGGTCCCGCAAACGCAAAACAAAAATCGCCAACATTT
 GAGATTATAAAATACAAAAATCTACAAGTTTGTGATGCCGTCTCCATCTCCTGATGCCACT
 GCCCTAATTAGGATTATTAGCGCGGAGCTGACGCCACATTAATAAGCTGCAAAATTTGTT
 CTCAAAATGTGTATTTCTCCTAAGTATTGTATCTTCAACGAATTTCCGCCAACCTGTAA
 CGAAAGGAAAAATTAACAACAATATATACAAAATTAATCAATGACACAGAAAAATAGCAAAC
 CTGACAGGCAAAGTATCTGATGCAAATAGGGCGACTCCATCCTGCCGACGACAAGCACGC
 AACTTCTTCTCCAAGCTTACAATTACTGAAATAAGGAAACACAAGCCAACCTCTCGGAAA
 TATGTATCTAGAATATAATACTTATCTAATTGCCAACTCAATGATTCCGAAAACGCGGGA
 TCCCAGCTGGAATGGCACAAAAATATAGGCCCTCCAATCAGCTCTCACAAAACGTGCCTG
 AAAAAATAAAAACTAATGCAAA

sd 42A

TTATAAATATTTTTGTGTTTATTTCCATCTTACTATTGTAGTAAGAAATATGTATATGA
 GTCCCAATATATTGTACATGAAATAATTGATACGAATGAGACCTACAGTAAAGTAGAATA
 AGAAGCAATGGAGGAATAGCCACAAATTTCCAAAATTCACAAAAATAATCATAAATACTTA
 CCTCCAAATTAGCTTCCACTTAATGTACCTGAAAAACAAAAACAAAAAATTAATAAAATT
 AAATAAATACACGAACAACTACGTGCAACGCCGCGCCGAACATACATGTTGCCATCGCTG
 GCGGAGGTACCTGAAAAGAAACCAATTAACAATGTTAGTTCTAAATTTCAATCTTTTTT
 GTAGAATAATTCAAATTGTTAAATGTAAACAAACCTTGCAATACGTTAATGTTACCAGTC
 CATGTTACTGTCTAAAAGCTAAGTTTACAAAAATACTAATTATATACTAACTCACCACGC
 CCAAGCACCAAACTCACCCCATGCAATATAATACTTATAAATTTAAATAATTGTACCTAT
 ATATTGCACACACTGTAATCAAAGGCAAAATAAATCGTGGATGCGGAACAGAATTCACCTC
 TGTCCTCGTACCTCCACCAGCAAAGTTAAAA

HeT-A 9

GCAAAGTTAA ATAATAAAAAATAAAATAAAT
 AAATACTTAAATTAAGTAATAAAAAATAAACGCGCTTCGTCTGCAAAATACTCTCATGCG
 TAAATTTAATTAATAATCGACTTTCAAGTTTATAAGTTAAAAGTTTAAATTTGTCTTCCG
 GCCGCAAAGTTTGAACCGCGACAATTTAAACATTTAATGGACAAAATAAAAGCGAAAAAT
 TAATCAGTGAATTATTTGTGCAAAATTGACAAGCTGCCGCCATAACCAAAAGGAGAAAGA
 AAAAAAGAGAGAAGAAAGCAGAAAAAAGTAAACCTAAAGAAGAAAAGGAGGAAAAACACAAT
 TAAAGAGGCGATCTAAACTGGAGGACACATTACAATTAAAAGCCAGGGTATTTATACCTT
 CACAAGTATCGTTTTAATATAAAAAAACCAACATGTCCGTGTCCGGCAACCTTTTCTCT
 GACGATTTGATCATCCAGAACAGCGATCTTCCCGTTCTACCTCAATATATATCGCCAC
 GTCCACGGATCAGACAATTCTCTGCTTAACACTATCATTTAATCAGAAAGAAATTGCC
 CTCAAAGCCAAAATAAAAGTTCAAAAACCTCTCTGCGGCTGCCATAAAAAATTGTTAATTCC
 CTTTCACATAAGAAGAAAGAGAACACAAACACTAATAATGCCCCAAAAAGACCCCTCTCA
 CTCACCAATTCTACTGCAAGCACTTGTGGCGCCAAAAGCAGAAATTCGAAGGGGAAATTG
 CCTTCTTCTCCGTTCTCCTCACACACATATGAGGGGAAATTACTCTCAAAACTGACTCAC
 ACTGATGCTGCTGCCACACTCACATCGACGCTAATCGAAGAGTGAAAAATATTCACCTCA
 GCGATCTGCACTATTGCAGCTGCAGCCACAAATACGGACACAGACTTCAAAAGCGCCAAA

ACGAGCGATGCTATGGGAAGTTTCCCCTCTCTCGCACAGTGACAATAGCATAGAGAAAAA
TCTTCCAATGCTTCTTCCCCTCCTTCTCATGCACACACACACTAGCAAAGCCACTGAC
ATAAGCTTAGAAAGCCGCTCAAAATGTCCCGCGCTTGCCAATACTGACACACGCTCCAAA
AAAGCCAATGTTAGTGACATTGGGGAAATATTCTCCCCACTTATACAAAACGCAAGCAAG
AGGACAGGCCCTTGACACAACCTATCAACGCTTTTTTGGTCGATTTTTTAAACCCCAACCCGGACG
TTACTAAACTAAGTCTAAATAGGAAACCCACTAATCCCCTAGAAAATACTGGGAAAAAAA
GCATTCCAACCTTTGTAACCTCCAGCATTTCCTCCGCTCATCATACCCGTTTTTCAAAAG
CAAGGCAGCTCAATCTGTTTATGAAATGGACCCCGCTCAAAGCCAGAAATGGACCCCC
CGCCAGCCGTTACCTCGCAGCATCAATGTCTCTGCTCGCAGCGCAGCGGCCACCCGGG
ATCGCCCCCTACCCCTCAAAACACAGATACAGAGCTGCCTCCATGGAAAAATCGTGCCC
CAGAGCCGTAGAGCGCCCCCTATACTCGTCAATGATGTAAAGGAAGCGCGCTTCGTGAGA
AACATGACGTATCCAGCTACAGGTGGACCCATGCGAATGTTTGAAGTGGAGATCATCATG
ACCAAGGATGGCAGCCATGAAAAAATTATCTCACTCAAACAAATCGGTGGGCAATAGGTG
GATATTGAAAGGAAAAACAGGACACGGTCCAGTGCTACAGATGCCAGGGCTTCAGGCATG
TTAATAATTTCATGCATGACGCCGCCAAGATGCATGAAATGCGCTGACGAACACCTGTCAT
CCTGTTGCACCAAGCCAATAACCAACCCCGCCACCTGCGTCAACTGCTCTGGGGACCATT
TTAGTGCATACAAAGGATGCCCGCTTACAAGGCCGAAAAACAAAAGCTGGCGGCCAACAA
ACATCGACGTAAGCAAAAAATAAGAACAAATCAAAGGCGCAACAAATAACTTTTATAGACGTC
AAGCCCCCACCCCCCACGCAATAACACCCCTCGGCTGCCACACAGCTCAGCAATCCTG
AGCAGATCAATCGCCGAAGCTCACCAAGAGGCAGCCAGAAAGTCGATGTTAAATCCTTTC
CGGCAAAATATGAACGACAGAAGACCACGTTTCTCAATGGCACAAATATATAGGGCCTCC
AATCAGCTCTCACAAAACGTGCCTGAAAAATAAAAACTAATGCAATCACAAAAATAAAAA
AATCATAAATTACTTCCACCTAAAGTACCTGAAAAACAAAAACAAAAATTAATACAATT
ATTAATCAAAATAAATACAAATATAATACTTACCTCAAAATTACCTCCCAGCTAATGTAC
CTGAAAAAACAAAAATTAATACAATATCAAAAAACAAATAACAAATGTAATACTTACCAAA
TTTTAATGTTGTATTCTCAATTTCCATGTTTCCAATCGTTGCGACTGTCTCGGCAACAAATC
GTGTTCCGGCGGCTCCAATGCTGTCAATCCCGACGCACCTGGCCACAAGACGCGGCGCTAC
TGGCAACTCTCGGTGAACAACCGAGCTACAATTTCTATGACGATAAGAGATACTCACCAA
ATGACTCCGGCGCGGGATGCTACCACCACGAACAACTACCTGCAACGCCGGCCGTACAT
ACATGTTGCAAGTGTCGCGCATCCAGCGCCGCAACATAGCCCCAGCCTAAGTACAACAA
CTACTTACCTGCAATGTCTCCAGAGGCTTCCAGCGACTCGGTGCTTCCGTCTTCTGGCG
AGGGTACCTGGAAAGAAACAAATTAACAATGTTAGACCTGAATTTCAATCTTTTTGTAA
AATAGTTCAAATTGTTAAATGTAAACAAACCTTGCAATACGTTAATGTTACCAGTCCATG
TTACTGTCTAAAAGTTAAGTTTACAAAAAACTAATTGTAACTAACTCACCACGCCCAA
GCCCCAAACTCGCCCCATGCAATGTAATACTCATAAATTCAAAAAAGCCAATGTTAGTGA
CAGTGGGGAAATATTCTCCCACTTATACAAATTGACGACCGCAAGCAAGGACAGGCC
TTGCACAACCTATCAACGCTTTTTTGGTCGATTTTTTAAACCAAGCCGGACGTTACAGAATT
CACTCTGAACCCGTACCTCCACCAGCAAAGTTAAAA

HeT-A 9

GCAAAGTTAA ATAATAAAATAAAA
TAAATAAAATACTTAAATTAAAGTAATAAAAAATAAACGCGCTTCGTCTGCAAAATACTCTC
ATTTATAGACGTCAAGGCCCCCTCCACGTAATAACACCCCTCGGCTGCCACACAGCTCA
GCAATCCTGAGCAGATCAATCGCCGAAGCTCGCCAAGAGGCAGCCAGAAAGTCGATGTTA
AATCCTTTCGGCAAAATATGAACGACAGAAGACCACGTTTCTCCTCCCATGATACGGCC
ATTGAGAAGCGACTGAATAAATGGCGCCGAACACCAATAAAATACCCAAAAAGGGTAGG
ATAGCTTCAAAGGACAAGGCAATGCCACGGCCTTTACCCAAGACAAGTAACCCAGCGCAA
AGACATCTGGAAAAATAGCAGGACATGCTCCGAAACGAAAGGAGTGAATAAACTGACCAT
GAACCTGAAAAAGGTACTTCTAATCCAGCCGAGTCGGCAACGACAGCCCTCCGACCACC
ACCAGAGCCGCCAGAGCTAACCTTAAGCCAAGAATTATCGACTAAGTACACCATCGCCA
AGAAATTCCAATCCTTACCTACAGAAAAGCTCGGACGACCCACCATAAACCTAGCTAAT
AGAGTCGACAATTTAGAAAAGTAAATTGACATTTTAAATGGCTTTAATCATACAAGGAAGC
AATAACAATAATCTTGACATCGATACATCCATCTAAATTTACATTACATTTATTTATATT
TATATCTATTATATCTATATATATATCCGACATAAAAGCGCACGTATGTCTACTTCCGTA
ATGCTCTACTTATCATCACCTTCTCGACGCATAGCCTAAACCACTGTTGAAAAACAAAT
CAATTAGATGGATGTCATAAAAAACGTAAATAAAAAATCCTCTCACCTCAAGCATCCGGAT

AAAAAAGGTAAGACAACGCACTCCAGCCGATTCTTGATGAAGCTATGTGAAGAAAACCAC
ACCAGGACTCAAAAGTCAAATCGGAGGATGGACATGAGAAGAATTAGTGCGGCAGAAGCA
TGATGAATAGAGGCGACTCGCTGCAGCAAAATATGCACAACGCCACTTACCTGAATCTTC
AGCGGCACAGTCTTTTTTATGTACCATTATCTCCGCCGCAACCCCTTCACACAGCTCCAC
TACATAAGACGTGCCACCTAAGCTGCCTCCGTACTGAATTGGACATCATGCGTTGCGCTG
CAAAATCCAATCCTATCGACGAGCGCCAACAGCGAGTCTGCGCTAAAAACCAAAAAACAA
AATAAAACAATTAATCAACAGCAAATTGAATATAACAATCAAACAATAACAATCACTTAC
CTCCTTAACTGCATCCAATCGCTGACTCAAATCCAACACAACCGACAACAGGAGACGGGT
CCCGCAAACGCAAAACAAAAATCGCCAACATTTGAGATTATAAATACAAAAATCTACAAGT
TTGTGATGCCATCCATCCATCTCCATCTCCTGATGCCACTGCCCTAATTAGGATTATTAG
CGCGGAGCTGAAGCCACATTAATAAGCTGCAAAATTTGTTCTCAAAATGTGTATTTCTCC
TAAGTATTGTATCTTCAACGAATTC

ANEXO 2. Análisis de la secuencia de una mitad del palíndromo, presente en el BACR07N15

Se han anotado los elementos transponibles presentes en la primera mitad de la secuencia del BACR07N15, así como las partes de la secuencia con duplicaciones segmentales. Los genes de origen que fueron duplicados (procedentes del cromosoma X) son *ade5* y *CG12717*. Sus nombres han sido escritos en verde. Por otra parte, los nombres de las regiones duplicadas, procedentes del cromosoma 2L (2L35 y 2L38) han sido marcadas en naranja. Se ha empleado un “-“ delante de los nombres, cuando las secuencias se encuentran en el sentido contrario al mostrado en la base de datos.

F

```
CACGTTGAATTCCTCAGGTGGCTGGGCCCCGTAGTGATGAAGACAGAGCCAACACATTTG
CCGCTCACCTACAAAATGTGTTACGCCAAACCAGGCTACTAGCACATTCGCGCTACCG
TCCTATCCCGTAAACCGCCATCAGCAACTCACCCCAATTGTGTTTCGTCTTAAAGAAAT
AACTAAAAATAATCAAAGACAATCTCAGCCCCGAAAAATCCACCGGCTGCGACCTTATAA
CACCGGAAATGATCATCCAGCTGCCACATTCTGCAGTACGCTACATAACCAAGCTCTTT
AATGCCATCACCAAACTTGTTACTTTCCACAACGATGGAAGATGATGAAGATCATAAT
GATTCCAAAGCCTGGTAAGAACCACACAGTCGCTTCATCTTACAGACCAATAAGTCTAC
TCTCATGCATTTGAAACTATTGAAAAATGCCTGCTGATCCGACTTAATCAACATCTG
ATATACCACAATATAATCCCAGCCACCAATTTGGATTTTCGCGAAAGCCATGAACCATT
GAACAGGTGAATCGTATTACAACGGAAATAAGAACTGCATTTGAATATCGGAATACTG
TACAGCAGTATTTTTAGACGTATCCCAAGCATTCGACAAAGTCTGGCTCGACGGCCTAA
TGTTTAAAAATTAAACATCCCTACCCGAAAGCACACACAACTTCTAAAGTCTTACCTC
TATGACAGAAAGTTTGAGTGCGGTGCAACACTGCCACTTCCACTGTATGAACAGGAGT
TCATACTGAGGCTGGAGTCCCCAAGGCAGCGTTCCTTGGGCCAACCTTATACCTCATCT
ATACAGCCGACATCCCTACAAATAGTCGCTTAACGGTATCCACATTTGCCGACGATACA
GCTATCCTTAGCCGTTCAAGGTCCCTATCCAAGCTACAGCACAGTTGGCACTGTACCT
CATCGACATTGAGAAGTGCTCTCTGACTGGCGAATAAAAGTAAACGAGCAAAAAATGCA
AGCAGCTGACGTTTACGCTAAACAGACAAGACTGTCTCCGCTCTTGTGTAACAGCATA
CCACTCCCAGAAAGCAGATGAGGTAGCGTACCTAGGAGTACACCTTGACAGAAGACTCAC
ATGGCGCAGGCACATTGAAGCCAAAAAACCCTTAAACTCAAAGCCAACAACCTTAC
ACTGGCTCATCAACTCTGGTTCTCCGCTCAGCCTAGATCACAAAGTCTTGCTCTACAAT
TCTATATTGAAACCAATCTGGACCTATGGCTCACAGTTATGGGGCAATGCCAGCAACAG
CAATATTGACATCATTCAGCGAGCACAAATCAAAGATTCTGAGAACCATCACTGGGGCAC
CGTGGTACGTTTCGGAGTGAAGAAATCCAAAGAGACTTAAATATCCCATCAGTTACCAAC
GCAATCACGGAACCTTAAGGAAAAATACCATAGCAAGCTTCACACGCACCCCAACCACCT
AGCGCGAGGTCTAATGCAGCTCAGCAGCCGTTCCCGTCTCCGCGGAAAGGACCTACCAA
CCCAGCGAATAAATTATTAGGGCCGTTTAAATCATAGAACAGTTGGAAAAATAATACAAC
TGTTCAAAAAATACTTGTTATAGTTAAGATTTTTAACTTATTGTTGGTTCTTATACAA
GAAGATTCAATAAATAAAAGCAAAGTAAAAAAAAAAAAAAAAAAAAA
```

ade5*, *CG12717

```
TTAAGATTCCCACATTAAATGGCCAACTGGTGATGGATGAGGATAGCGAGGAGGAAAAAT
TTGAAAGAAGAGCAAAGAAGAGAACCTGCGCAATGCAATGTAATTGTCGTAATAAGGC
AGCAATTTGAGCAGTTAATAGATGAATAAATTAATATCAAGTAAGATTGTATTACTT
CTAGTATAATATATATCTTACAGTAAATCAAAGAGACGAACATGTTGGCAATCATATGA
```

AACGAATTTACGCATAGCATCAGAACCATCGCCCTGGCGAAAATCTTGATTTC AATTAT
AGATTTCTGAAGGCTTTGAGTTCATAAACGATCAATCTTATCAGCGATTGTTGACGCTGC
TGCAGTGCATCGATAGCATTTGAATTTTTTTTACC GCAAGGGGAGACAGATAACACACAG
ATCGCAACAAATCGAATGGGGTGATAACCCTCAGCTTGCCATTGGCTGTTTACCCACCC
ATACCAACCTGTAGCCCTTGGGCACGCCGACATTGCGATTGAGGAACCGGTGGCCAGGC
GACGGGTACCCACTCAATGGGTATCATCTAGCACTTGCGGGCGATGAAGGACTTGGCG
CCGCACTGCTTCACATAGGCGGTGCGAATGAAAATTAAAAAATTAGAAATTGACTGAC
GTGCATTATAGTTTCATTTGACTTATAGTTTAGTAACCTCACCTGCCTCGTTGAGCAGC
CGGACGACCTGACCGTTGGTGGTATTGGAGATCTCCGCCTTGCCGCCAGATCGTGTC
TTTGACACCATCACCGGCGGTGATGCGATCCTTGCTGAGGAGCAGGCAAAAGTCTTGCT
GCTCTGGCAGATCGTACACCTGCTTGGTTTTTGCCCTCGATGATAACCTTGCCGAGTTTA
TAGCCCTCAACGATGCTATATGATATATGATGATACTGATTTAACGAATGGAATTTGTT
TACGAAATCCACTATTATCTTGAGTTAAACAATGGAAAGCCGCCCTTCTTGTTTACAGC
AGACAGCTCAATAAACAGCATTAAAAAATGATAGTACTTTATTCTTGATTATCAGAACC
AGGAAAATATTCCGTTGTTTTTTTTTAAAGTAACCCGAATAACTGTTTTAATAAATCCAA
GGCTGGCGGGAAAAACATAAACAACTGTCTTGCTTTCCACCACATTAAGATTGTCTTTTA
ATGATAAAAGATCGGATAAGCTTTAAGGTAGCTTTGGTAATATTACGTTCTAGGCACA
ACAAATGTTTATTGACTTGCTGGGCTTTATGAAGTTTATAAATTTATGTTTTCGGGGAA
ACCTTGGAATGGATAAACTTACTGGACGCGGTGCTGGTGGTGGACATCTCGTTAATCG
CACGTTGTTGGTTGCTTCACTTTAAGAATCGAAATTGAAATCAATTGGCGGCTGTCTGT
TTCACGTGTTGCGGGCTTTGCTTTATGCTCGATATTCGCTGGTGCCTTCCCCCTCGCTC
TCTATCTTTTTCGCACTCTCCGCTTGGCTCTATTATTACAATCCGCAAACTCCGACGTT
CGTCTGCATCAAAGAATCGCTCCAAAGCAACTGAGCGTCTATAGAATCGCCGGCGCGAC
TTTATGGCTCGATTTTTTTTTATCGCGGAACCCCAACACCAATGCAAGAGATCCATGTG
TTTGCGCTCGCGTGTGTATGTTTGCCTGTGTTTGTGTATGCTTAGAGTCGAGTATAATG
CTGATTGGTAATACCTTGCAATTTCCCAAATGGCTGATGAACACGGCGCGGGGCCATTT
AATATGTACTTATGTACATACATACAGTTTTTGGCAATCTATAAAGCTTTTCTTAT
GCGTGACAAAATTTTATTGTAGCATTTTATTTACAAATTTTGTCTAAAAATATTTTCTA
TTCATTTGGTGACTTTCTAAATTTTGAATTTTTTTTTGATTTTTTGGCCATTCTTCAACTC
CTGACATCCAGCTCATCTGCCTATCATGCTCATCTCTAGTAGGTAGCCTCTTTATCG
TCTTGTGGCCGATGTCACTATGTCTCAGTAAAGGATCGTATCTGAATGATATATTCAT
TGATTTTCATCTGCGCTGGCTAAAGAATAACATCAGCTGGCTTAAGACGCACATCTTTA
GCACATTTTCCACAAGAGATTGACCACACGAACGAATCCACAAATATGAAGCTGACG
GCAGCCGAGAAGCGTCACGAAAGGGTTGAGGAGGGGACACGAAACGTGAATATATTCTGA
CAAGGAGGACTTCATCATAATACCATTCAACGAGCAGTCCCATTGGATACTGGCCATCA
TATGTTATCCAAACCTTGGAATCCCGTCGATACACAATAAATGTACAGACGCACACTC
TGCGATGATATTCCAATCAAACAGCATTGAGTCTTATTTTCGATTCTGTTGGCGGTCACC
TTACGACATCGAGCGATTGCCATATTGCGGGATTGTCTCACCTGTACAAGGC AAAAGTAT
CCGAATGCGCTGGCGCATATCTTCAACAAGGATAATATGCCCGCTCATAGAGTAGAGGT
GCCGCGGCAGCAGAATCTCAGATTGCGGCCTCAGTATGTGGAGCAATCTTTACAAAAC
CCTTCATCGACTATACACTGCCATTGGGGAGCTGAGCAATTGGTTTGACCTGCTCACA
GTAACATAAGAAGCGTGAGGATATCGCTAATCTCATCCAGAAGCTAATGAATGAAAGCAA
TCAGCAGCGCAAAATTTTGGCGGTCAATTAAGTTTCCACATTAAATGGCCAAC TAGTGA
TGGATGAGGATAGCGAGGAGGAGAATTGTAAAGAAGAGTATATACCAGGCCACCCGTAC
CGGTTGGCGAATATGTGAAAAGGGTTTCTCGTCGGGTGAGAGTCTTCTCTGTTTGGC
CATCGTCGGCGGCGGTGGTCTCCTCATCGACGCTGCTTACCAGCACCACGGCATCTTCA
TTAACGTCCCTGCTTAGTATCCACTTGCGATTCCGGCGCAGGTGACTCCGCTCTGCGCG
TTCTGCTTGTGTGGCGAGGATATGGATGCAACATTGATTTGCACCCGCGATTGAGATCC
ACCAGCTGTCTCTTCCGATTTATTGCTAACAGAGGCATGTGCGGCATGGTTCTTCGACT
GGATCCACTGGGAGATTGCGGGAGCATCTCCAGCTGGTTGCATGCCCTGAAGACCATTA
CTCGTCATTAAAGAATTGGAGGTTATAGCCTAGAGTTGTGACTTCCCTGTTTATTGCCA
AATGAACCTTGGTTGTCTTGAGAACCAGCAGCTTGATAGCCATTGTTATCCTTTGCAAA
ATGACACTTCTTCTGTTTCTTGTCTTGGTATTATTATTGGCACTTGCCAGCATTTGGT
GACTTCTTACCCAGGATATCAGTTGGCCAGCATTTGGTGTCTCCCGAGAGCAAGGAC
TTTTCTTGTGCTTCTTGCCTTTGGTTGCACTTGACAGCTGGCTGGCACTATTGTTGGCC
ACCAAAGTCTCTGATGGTTTGACAACCGTTGGAGAGAGCTCTTCACGTACCGCTTGGTC
CTTCAAAAAGCTACTTTGCGATCCATTTTGGATTTGGATTCAACAGTACAACATATGC

ACGTTATGGGCTAATCCGACTGAGATGAAGTTTCTCGTTTCGATTGCAAATTCCTCGGGC
 TGCTTAATGTGCAAAGTTAACTGGCTGGTTGAAGAGTCACCTTACACGGTGACTTTGAT
 CACTATATTCCTTTCATTTCCCAATCTGTGTGTTTTATGCGATATTTAATATCTCTCTA
 TTTTATATTGTTGGGGTAAACGGCGTTCCAAGTGCTGTATCCAAAAATCATCTGTGCGAG
 TTGCTATCGCCATCGGAGTTCTGCTGCTCATATCGATACATATATGTTTCCAATTGATA
 TGGATTTTGTGCGTTGTTTTTAAAAATATGTATCGATGTCCGCTTTGAAATTTGAATGT
 TGGGAATATTGAAAAATAAATAATTTGAACTGACCATTGATCATTTACAAAATTAGT
 ATTCTTACATAAGGTATTTAGAAAGTTAAAGTTTACGCCGTTTGTGCAAGTCTCTTGTC
 TGGAAATTTTCAGAAACAAAAATGTATTCTCGTATTAATTTTCAGCATTTTTCGCAAAATACAA
 GTACAGGCATTAATAAAATTATATTTGTGCGTCCGTGTTATCGGGTCCACTTGTCCGAT
 TTTTGTGCTCAGCGTTAAGCTCCTCTCACCACATGGATTGCTTGGCGGCGATAAGAAAA
 TCAAATGTAAACAAAAATATTTTTGTTTTTTTGCCAAGAGCGCTTTATAAAAAAAACT
 GACAGACACACACGTTTTCTGTTTCGCGGGCCAAAGGGTCCATTTTCGCAATTTGCA
 TCGGTTGCCAAGTGCGGAAAAACGCATCTAAATCACAACGTTGTAAGCTGTGAGAAAGT
 TTGCCAAGTCGAATTCCTTAAATGTGAAGGAAC TGAAACTGGCTTTTGTTTTACATTT
 ACAAATTTACTATTACTACTACTATAT

- Idefix

AATTAAATTCCTGCAATATTTATAAGTGAAAAAATATAATTTTTATTTCACTAAATGA
 TTACAAAACCTTTTTATTTGAGAAGAACACCGACCTTTTACAATTAGGTCTCAAAATGAA
 CTCTGATAATTATTCTGATTGATTGACGTGTATTGTAAGTGATTAGTATGTAGGCAT
 ATGCTTAAGTCTTAGGGACTTGTGGATATGTCACCTCCCCAACCTTTGAATTTGGCCTG
 TCCTCAGGTGCTTAAATTTGGGTATAAGGTAATATTATTTCTTGTGTAATATTCTCAT
 TGACATTATTTTTCAAGGTGCGTGTGTTCTTTTTGCTTACGATATTTGACAATAATTTTC
 TTAGGTATAATTTTAAACGTATATGCTACATACAGAGTCAAAC TAATTATCATTATGAC
 GAACGTTGTAATGCATATAATTTGGAGAAAGTTATAGTGTTTTACATGGGATTCTATCA
 TTTCTCTCGTTTGTAAGTGATTTATTGGTTCTAGACGTGTTACATTGTTTGTGATATAA
 ATTGTTTGTGTAATACTGCTATATAATAATATTAGAGATTAGAGATTAGAAATTCATC
 TATTTGCGATGGTAGAATTTTTAATTTTAATGATGTTGTTTCTGATACTAATATTTTAT
 TGTTTGTACAGTCTTAACAGCTGTTTCAGGAATATTCATGTTAATAGTATATTTGGTT
 CTATATAATTTATTTTGTAAATTTTTATGTGTTTTAATATATTTGCATTGAGTTTGCTCT
 TGTTTGATGATTCCAATAATACATTTGTGCAATATTACCCTATTGGTATCTTTATGGAA
 TACTTGGTTATCAAATATGTAGAATTTATCGAATATTTGCTCGGTTAGAATATAATTAT
 GTTCATCTGGGTACGGAACAATTTGGAATATTGGAATTTTAGTGACTTCGCTACATGGG
 AAATAATCAAAATTTTCGTTTCGTATCTGTTTTAAGCCAAGTTGACGTTTTTATCTTTAGC
 ATTTTTTCTGAATTTACATGTTTTAAATAGTCATGCTTTAGTAACTTTGATTGAAAAAT
 TCCAAGTCTGGTTAATTGCATACCCAATTCTATGCTTCTATATATTTCTGTAATTTGCT
 GTAGGTTAAATATTAGTATTGCACCTTTGTTGGTGTGTTCTTTATCTACTTTGAGCTTA
 TTAATTATATCTATACCACTATTGATTATATCTATAATCGTGTTTAGGACATGGGTTTT
 TACATAGTCTTCTGACATGTTGTTTTATTTTTCTTCCAACCTCTCTCTGTCGCTTTCAT
 CTAATGTGCCGAATAAGAATTTGTATGCTTTTCTTACTACGTTTAAAAGACTTCTTTTA
 CTTCGACTAATAATTTCTTAACCAATTTATTTCTCGTTTTAATTTGTCGACTAAATATTG
 TATTTGTGGTACATTAGGATAGTTGTTTGTCTTCATTTACTATATCTTCGAACATGAGCA
 TTGCTCTTGTATTTTTCCATTGTTTCTGTTTGGAAATATAAGATATCCATTCTTCGCGC
 TTATTGGATTACTTCTATATTGTTGCACCTTGACCATGGTGAAAAGTGATAGTATGCAA
 CTAAGCATTATTAGAAGTATGTTGTATACCTTTGTGCTTCTTTGGTTAGTCTGGAATTAT
 CATATTTGCTCTTATTACATTTTTTCTGTTTTCTTAAATTTTGTCCCTATAGTGAACGATT
 TTCTTACCACGGTTTGTCTTTCGAAATGTTTATCATCTACTTGTCCAATCCTTCTGT
 ATTTTTGAATGGATTGGTTATTTTTACTTTTACTTCAAAATCGTGCCTATCTTATTGAG
 GTATTCTATCTTATCGATTTTATATTGTTGTACGTTAAAGTCTGGGCTGCCTGCATATA
 GGAAAATGTCTGCTGAAAATCTTTTGGTACTATTATGTTTAGTTTTATGATTCTATATG
 TATAGAATTGTTTCGAATTTTGTGCACCTATCTTCAATATTTTGTCTCGCTACCAATGAT
 TCTTAGTTTTTTCGTTTACGGTCTTGTGGAATCTATCTATGTCAGATATACCATTTTTGC
 TAGTCGTTATAGAAATTTGTACACCTTCAGATCTTAACCAATTTTGTAGGGCTATACAC
 ATAAAAGCTGAGTCTTTATCTGCTTTAATTTCTTTTCTTATATCATTTGAATATTTTAGT
 AATGGCTCTTTTTGCTTCTAGCCAATCTCTACTTTTTACTTCAACTAGTGATGCAAATT

TCGAATAGATATCAATGCAAGATAAGAAGATCTGGTTTCCTGTGAGATAAAAAATCTATC
 ACGTATTTTTCTCTGTGTAAATATTTCCGGTGTGTTCATATGTCAATTCGTGTT
 TCGATGTTCTGTTTTTCCAGATTACAAATTTACATTAACTGATAATGTTTTGAATAA
 GCTTTTGACTATCAGGATAGTAATATTCTTTTTTAAATAAAATTGATTGTTTTTCTATA
 CCCGGATGTAAAAGTTCTTATGTTTTCTTTAAGATTAAATCTTTAAATTTCTGCATACGT
 TATTATATCAATAAACTTTGTCTTAAGAGTTTTATGAAATTGTTAGGGCTTATAATTTT
 GGTAAACGCTCTCTGGAAGATCAGAAAATCTTCGTCACTACGGAATAAATTGCACTCT
 TTTTGGTACACACATATTCTTAATGAGGTATTTGGTTCAGGTGTCATTTCTTTATAGA
 CTATTTTAATCTTTAATCTTTGAAAGAAATGTTGTACACTTGTGTATCGTTGTCTGACT
 TTTTCTATCTCTATTTGTCTAGAGAAGAGATTTATTGGCCTTTCTGTTATAGATATGTA
 ATTTTGATTATCTTCACTGGCGCTGTGTATAGTTGCGTCTGCGATGTTTTGCACTCGC
 CAACCATAACTTCTCCATTTTCGTGCGGAAAGAGCATCCGCGACATGATTTCTTTG
 CCTGGAAGATATTTGATTTTGTAAATCAAATACATTGAGTTTTATTTCATCTTTGCAA
 TTTCATATTTGGTTCTTTAATGTTGTTGAGTCATACCAGTGGCTTGTTGATCACTTAATA
 TTTCAAATGGCCTGTGCAATAGGTATGACCTGAAATATTTTGTAGCCCAAATTTAGCT
 AACAAATCTTTTTCAATCATAGCATAGTTGATTTTCATGTTTCGTTTAGCGTTCTACTGGC
 ATAACAACTGGCTTGATTTTATGATAACACTGCAACATTAGCTACATTACTGGCAT
 CAGTTGTGACAGAGAAAAGGGTTTTGAAAAATCAGGATAGATTAAATATCGGGTCTGAAGTT
 ATCAAAACCTTTAGTCTTTCAAATGATTCGATGTAGTCTTTACATTTGGTGTCTATTAT
 AGCACCTTTCTTTAATTTGAGGGTCATGGGTTTAACTATTTTAAACAAAGTTAGGAATAA
 ACTGGCGATAGAATCCACATAATCCCAAAAATGATTTTATTTGCTTAGGTGTCTTGGGT
 ACTGGAATAATTTGTAATTGCTTTGGTTTTATTTGGATTTGGCTTGATGCCATTTGTTGT
 GACGATGTGTCTTAGGAATTCAGTTTCTTTCTTCATGAATTCACATTTATATAGTTGCA
 ACTTTAAATTAGCGTCTCCAGTTTTTCAAAGACTTTCTTTAGGGATAAAATGTGTTCT
 TCCAATGAAGTGGAATAAACAATAATATCGTCTAAATAGACTAAACAGTCTTTGTAGAT
 TAAATCTTCCAGAAGATTATTCATGCATCTCTGAAAAGTAGCTGGAGCGTTTTTTAAAC
 CAAAAGGCATACGAGTATATTCATAATGCCATGCTTAGTTGAAAAGCTGTTTTTTGCA
 ATAGAATTTTCATCCATTTGGATTTGGTGAAATCCCTTAGCTAGATCTATAGTTGTATA
 GTATTGGCATCTACCTAATTTGTACAATATCTCATCCATTCGGGGAATGGGAAATTTGT
 CGTATAGCATATCCTTCATTTATTTTATTGGATTCTACCGACTGCGCCAATTAATTTCT
 TGCAATATTTATAAGTGAAAAAATATAATTTTTATTTCACTAAATGATTACAAAAATT
 TTTATTTGAGAAGAACACCGACCTTTTACAATTAGGTCTCAAACCTGAACCTGTAATAATT
 ATTCTGATTTGATTGACGTTCAAGCCTCGGTATCGAGCTTTTCCAATATGGACTTTAGA
 CTTTAGGATTCTGATCGTGACGAGAAAAAGCTTTGCGTTTCTCTTGGATTCAAGTGCAT
 TTGCTCTTGGATACAAGTGCTCTTAGATTCTTATAATTTGTTTATTGAAATCGATACTT
 TTGTTTTTCGGGATTGCGGTTCCGACCTTTGAACGTGGGAACGTGACGATGGGGTGAGG
 CTTAAGCAAGGAATTGTTTAAATTAGGGGACGGATGTTTAGTCTACTAGTGGGTGACTT
 ATCTAGAATTGGGGTTTTTTCCATTCTCGAGGATCATATAACAATCTTATTCTTATCTG
 GTATCTGTTGGGTACATCCGGAGTAGTGATGTTTTATTAGTGT

invader1

GTAGGCATATGCTTAAGTCTTAGGGACTTATGGATATGTCACTATATATACTTTACTAA
 ATAGTGCTTATCTGTGCTGGCTGCAATAATATATTTTGGAAAAATATATTTTTATTCCG
 GTTATTTCTTCTATAAATATATTAATTATATATGTCGCAGATCATTTAAATGTAACGAC
 ATTTCTAGTAATTGCGCGTTTCTGTTTGGCACGCGCCAATTCCTTTGATGCGCGGCAGA
 GAACCGTTAGAGTTTTAACGGAAGCGTACAGAAAAATGCGGTGTGCATTAGTATTGATG
 TGCGTGGACTGGAGTGGGTCTCCAAAATGTGTTGGCAATGCGTGTAGAGTTAAGAACTA
 CGCGTAATTCGGATTTTGTGAAAAGAGATTGCATCGTTGAGCAGAACAGACATGCCTCG
 CTCACGCTGCAGAAAACGTTATGCAACAACGAGAATAGTGAAGAGGAAAAAGAGTGTCC
 TGATGACTCTTCGACATCAGAAGTGGGATCGGGTCTTATAACTTCAACAGTAGATGGAC
 CATGGCGTACTATTATGGAATGCAAAATACAAATTTAATTGAACCTTGTCGGGGCCATA
 CAAACAACGGCATCTGATGGTAGCAAAACATTAACCGAGAAAAGCGAGTGGAGCAGTGC
 AGTGTTCGATTACCTAAGGGATTCTATAATCATAGAGGCAAAATTTTCGAAATCACGTT
 TGATTCTGGAGCTGAATGTTGCTGCTGAAAGAAAAGCTAAGTATTAAATTTTCTGGTA
 AAAGATTTAATAATGTTGTTATGTTAAAGGCATAGGCAATAATGGAATATTTAGTACT
 GTACAGGTTTTAAGTAACGTAAAAATTGACGATTACTGCATTGAAATTTTGTATATGT

AATAGGCGATGATATTGTTATTGGTCGCGAAATATTAAATTTGGGTTTTGATATTGTCA
 TTTCGTCAGACCGATTTAAAATTTTCGAGAACC AAAATAATTAATTTGTGTACCGGCTGT
 GATTCAGCCGATATTGATACTGAACTTGATGAACAAGATAAAAATAAAATTCGAATTCTC
 TTTTAAAGGGTATCCCATGTACTAAGGTTAAAGTTGGCAAAATGAAAATAAGATTAATT
 GATCCAAGAAAAGACAGTTCAAAGAAGGCCATATCGATTAAAGCCCATGCGAGCGAGATTT
 AGTTCGAGAAAAATTGACGAATTGCTGAAATGCAATATAATTAGACCAAGCTGTTCCACC
 TTATGCAAGCCCTATATAGCTGGTAAAGAAAAAAATGGTTCGGATAGGCTTTGTGTGCG
 ACTATAGGGAGTTAAATTCAAACACAGTCGCTGATAAATACCCTCTGCCATTTGTACAG
 GATCAAATAAGTAGACTTCGAGGAGCTTATTATTTTACATGTTTAGATATGGCTAGTGG
 TTATTACCAAATTCCTTAAATGCGGATTCAATTGAGTATACAGCGTTTGTGACACCCG
 ATGGTCAGTATGAATTTTTATCTATGCCTTTTCGGGCTTAAAAATGCGCCGTCGGTATTT
 CAACGATTAGTAATGAAAGTGCTCGGGGATCTTGCAAATTCCTATGTTATCGTCTATAT
 GGACGGCATCATGATAACTGCATCAACACAGTGAGAAACGTTAGAAAGATTACAAATTA
 TTCTAGATGTTCTGACAGAGGCCGGGTTTTTCATTCAATATTACGAAGTGTTCCCTTCTTG
 AAAACATCTATTCAATATTTAGGCTACAATGTTTGTGCTGGCGAAATTCGCCCCAAACCC
 TCAGAAAAATGTTGCGTTGACTGCTCTTCCGCCTCCAAAGTCAGTCACATCGCTTAGGC
 AATTTATTGGATTAGCAACTTATTTCAAACAGTTTATTAAAGGGATTTTCACAGTTGATG
 AAACCGTTACACTTTTTTAACGTGACACAAAAATAAATTTATTGGAAGCGGAACATGA
 ACAGATACAGAAAAAGGTTATTTATACTCTAACTGATAAACAGTCCTTGTATTATTTTG
 ACCCAAAATATCCAATTGAGTTGCATACGGGTGCAAGGTGCAGTTGGCAATCAATTGTA
 CAACTAACCGTGTGACGAAGGCAAGTCTCTTGAAATGTTATTGGGAAAGGTAACAAGC
 CTATTAGGATTACTTCCACCTGGTGACATAGAAAATGATGTTGATTGATGTTGATGTTTC
 AAGGGGCCTTTTTGAAATAATACAGGTTCTGGAGGGTGACAGGTACACTTTAAAGTCCTT
 AACCAATAAAAGATTTTATAAATATGCACAGGAAGATTTGAGAAAAATGCCAGAAGGGC
 AAGTGCCCCGAGGAGATGAATGTGTGTGAAGAGAGCGAAAACGAAAATGTTGAAGCCATT
 GAAGAAAGTGAAGCACACCAATAAAGGTTATGAATGATTTACTGCTGGCCAAAGAAAAAT
 GGCAGAGTAACAAGACCAGACAAAATAATGCAAGCACAAGTGCAAAACAAACTGATGTTT
 GTTTTCATCATGGCGTGTGGAGCATCGAATTTGTAATGATGAATGAAGAAGAATAATTG
 ATTTTGTATAAAATTAACATTATAGTAAGAATTATGTAAAGATATTGGTTAAGAAGGCTT
 GATACTGTTTTCTAAAATTAACTTGATAAGTTTATACGAGTTGAAATGGAATATTGAA
 TCTTTTGTGTGTAAGGAAAATGAATAACTAATAGTAAATTTATAAAAGACAATTGA
 ATGGAACACGACCTCGTGTTAGAATCAGGATGGCCGAGTCTGGTCTCCGTTTTGTGTC
 AGTTAAGTTTAAATCTGGTAACGGTTGGTGGTGTGAGTTAACTGGCCCGTATGTTGGCC
 GGATCGTCGAGTTGTAAGCTGTGAGCTAGGCTAACAAAAGACTGGCCGGTTTGAAAAAC
 TGTGCGCTAATGAAAAATTTTATTGGAATCCAGACCTTAAACGATTTATGTGTGAAG
 AATATTAAATTATGATTGATTGCTTGAGAGTTGTTGGTAATCCGGATTTTCATAGCTTGA
 ATGTAATAGGTAATATGTTTTGAATATTTGTATTGGTTTGCCTTGAAGTCATGAAGAAT
 GAAGAATAATTTAACGAATTAATAACTATGATCGATATTAAGGGATTACTAGCACACGA
 GGACGTGTGATAGGTCAGGATGGCCGTGTCGAGATCATTTAAATGTAACGACATTTCT
 AGTAATTGCGCGTTTCTGCTTGGCACGCGCCATGCAGGCGCCTTTCTATTCTTTTGAT
 GCGCGGCAGAGAACCTTTAGAGTTTCTGCCGAACGTAGTCTGGTCTCGCGAGTAGGAGCGG
 GGGGAAGTAGATATCTGCAGGAATAACGGAAGCGTACAGAAAAATACGGTGTGCGTTAG
 TATTGATGTGCGTGGACTGGAGTGGGTCTCCAAATGTGTTGGCAATGCGTGTAGAGTT
 AAGAACTACGCGTAATTCGGATTTTGTGAAGA

ade5, CG12717, -2L35B, 2L38

GAGATTGGGTCAGTTAACGTGAGAAAACGTTATGCAACTGATGACCTGATGACTCTTTG
 ACATATATAATATTTAAAATACATTTAACATTTAAGCATGCTGAAAAATGCATATGTAAA
 TAATACAATTAGCCTGCTCTATTTTTTAAATATCCAGATGAGTACATCATATTTGAATGA
 ATTTTTCGTAGTAGTCATATAAACTTTTTTTTGTCTCTAGCCAAATCGTTCGCATAGAAG
 ATGTCATATTTAACTAACAGCTTTATTTATTTTTTTTTTAAATACGCGTGAATAAAAGAGA
 AGAAAACGAATCAAAAATAACTGCATACGAAATAGATCGAAATAAAAAACGTAGTCGCTTT
 TATTCAAAACCCACAATCATTACAATCAGCAAACCTTCGACATTCGTCTGCATCAAAGAA
 TCGCTCCAAAGCAACTGAGCGTCTATAGAATCGCCGGCGCGACTTTAAGGCTCGATTTT
 CTATCGCGGAACCCCAACACCAATGCAAAGAGATACATGTGTTGCGTGTGTGTGTGT
 GGGCTTAAAAATCGAGTATAATGCTGATTGGTAATACCCTGCAATTTCCCAAATGGCTGA

TGAACACTGCGAGGGCAACTTAGGTACTTATGTACATACATATGTACATTTTTTGGAAAT
 CTATAAAGCTTTGCTTATGCGTGACAAATTTTATTGTGCGATTTAATTTACAAATTTTG
 TTTCTAAAATATTTTCTATTCTTTGGTGACTTTCAAATTTTGAATTTTCTTGATTT
 TTGGCCATTCTTCAACTCCTGACATCCAGCTCATCCTGCCTATCATGCTCATCCTCTAG
 TAGGTAGCCTCTTTATCGTCTTGTGGCCGATGTCACCATGTCCTCAGTAAAAGGATCGTA
 TCTGAATGATATATTCATTGATTTTCATCTGCGATGGCTAAAGAATAACATCAGCTGGC
 TAAAGACGCACATCTTTAGCACATTTTCCACAAGAGATTGACCACACGAACGAATCCC
 ACAAATATGAAGCTGACGCGAGCCGAGCGTCACGAAAGGGTTGGGGAGTGGACACG
 AAACGTGAATATATTTAGCAACAAGACTTCATCATAATACCATTCAACGAGCAATCCCATT
 GGATACTGGCCATCATATGTTATCCAAAGCTTAGAATCCCGTCGGTATACAATAATAAT
 GTACAGACGACACTCTGCGATGATATTCCAATCAAACAGCCATTGATTCTTATTTACGA
 TTTGCTGGCGGTACCTCACGACATCGAGCGATTGCCATATTGCGGGATTATCTCACCT
 GTGATACAAGGCAAAGTATCCGAATGCGCTGGCGCATATCTTCAACAAGGATAATATGC
 CCGCTCATGGAGTAGAGGTGCCGACGAGCATCTGGATGAGATTAGCGATATCCAATTC
 AGGCGACAATGAAAAAGCAAAAACACTATATAAACAATGTGGAACGTTTATAAGAAACA
 CATTATTTTGTATTGTATTTGAACACATTTGATCAATTTCTTGCTAATTTCCCGTAAAT
 TGCGCATCGACAGTTACCGCAGTTACCGTAAAAGAAAATCTCGAATTAATGCCGAGGT
 TGTCTTATTTGTGAAACTGGGGTGTAGACACAGATGGAAGAAAATTTGAAAAGAACTTT
 TAATCACTTTTGAATGCCCTAATTACTTCTCCTTGGTTTTCTCTGAAATCCTCACCAAA
 TATTAACGTTTACACTTACACGTATTTTTTTTCGATATTTTCTTTTAAAGTTTGGTAAAT
 TAAAAACGTGCTTTAAGAAATAAATAAATAAATCAGTCAAAATAATATTTTGTCAAGA
 ACCGTAATCAAATTTCAATTTGGTGATTCAAGAAAACCTGATAATAGTAGGGGAGCCGGG
 TGAACCCCGCCCCCTTACTTTGAACGACATAAAGAACGTGGTTCACGCCGAATTTGGGA
 TATTGAACCTCGTTTTGCTTAATAAAGAAGTGGTTTATGTCATACCGGACGAAGATCTG
 CTCCAGTACCTTGGCCTGATGACATGTTGCTGTAATTTTGTACAGAGGTGAGGGAATT
 CCTCCTGGATAACGGTGTGACGAGCTGGTAGCGCCTAATGTGCCTTCCCCGGCAGCAA
 CAAATTGAAAAACGACGGCCACACGTACTAGAGGGATGCATAACACCATTCTCCAGTAT
 CGATTTCGTTTTACTGCGTATTACAGAATACATGCCATTTCAGTTGGTTCGCGTTCTCTTTCG
 GTTCACTTCGTATGTATTTCGGACTTACAGGCGTCCGATCGAAAATAAACTAACTGAATGT
 GTTCAAAGAATGAATGAAGCGAATGAATTTTCAATGGTAATTCAGAGTACTCTGAATTC
 TTCATGTTGTCATTGAGTAAAATGAGTTCGGACAGCGCGGAGGTAAGTCAAAGTTTGTG
 TTTTTTATGTTTATTTGTATTATTATTTGCGTGTGGGTCTTATTTAATAATATTTGTTT
 GTTTAATTGTTTAAATAAATATTTAGCGATCTTTACAAATTGTGTCAATTATAAATA
 TTTTAAATATAATAGTAATATAAATAAAGAATCGTTTGTACAGCCTTTGTGAGGCCAC
 GCAAGTTAGAAGGTTAAAACAAGTAGTTATACACCATACTCGTAGAGTGAAAGTATGTT
 GTAGAGTGTAGTATAACTCTTTGAAAAGTTTGTAAACAATATCCATAATATTATCTCTTA
 AGCTACGCATAAAAAATATTACAAATAAAATTCTTCTTCAGTTTCAGTTTTTATAAAAACT
 AAATTCAGAGGGTAAGTGAATAAAACGGATTACGATTTATTTTACAACTAACGTTA
 AAAAGTGCAATTGTGTACACCAAAATTTCAATTATTTTAGACCAACCGCATAATTATTT
 TCCATTGGGGTGCTTGTCTGCCAATCAGTTTAACTTTTCCATCACCTTACGTTTAT
 GATTCTCTGGTTAAGAAGCTTTTTTTGAAATTCAGTTCTGTAAATTCATTGTGAGTCT
 AAGCGGAAAAATCGGCGTCCACAGTTTATCGAATGGACGTGGCCCTCCTATCTGGCTTTA
 AGCTAATGCCGGGGCAGTCCATTGACAGCCAACCACCTTGCCACGCCCCACACTCCAC
 ACCTTGTAATGACGACATTCACATAATGCAACAACAATTGCAGCAAACGCGGGCACTTA
 AATAAATGATAGTGCACGCGGTTTGGCCGAATGAAATATCGAGTTATTTCTGAAGAGGAC
 CTTCCTGGTTTCATCAGTCCATTAAAGATTTCAAATTAATATAGTATTTTAAATTAATGTT
 TGTTAAGCCCAAAATATCAGCTGGGTGCCATATTAACTCGGATTTCAGGAAACTTGAGT
 AATCTTGAATTATACCTAATATGGGCTTAGCTTTTTTATTTCTTGTAGAGGCCTTCTATT
 AATTATTTACCACTGCCCATTGCTGCCGTGTACTAACTTTCAAAAACCACTGTAGATT
 TCCGAGAAAAGTTAATAACCACTTCTGAATACTCATTGGTCTTAGCTAAAAAGCACCAC
 TTAGGCTCCACAGATGATGACTAAACGTCTGTTGTGCGGATGTTTGACCTGATATTA
 ATGATGGGCTTTATTGCGATTCCATCCTGGTTGCATAGCAGGCTTGGGCTTTACTTCT
 GGTTTGTAGACTCGGATTCCAAACTTATTCTCTAGTACGCAACTCTGCGTATTTAGGGCC
 AACATTCACATGCGGAAAAACATACTCCATCACAGTATGCTTACATCGTGCACATCAAA
 TAAATATCGTATTAACCTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT
 ATTTTCAGTCCGATGACAAGATCCCTGCGCCGAGCTTGTGCTAAGCCCAAGTATGTGCC
 AAGCAAAAAGCCGAGAAAGCTAAAAGGTGAAGTGACCAAACTTGACCAATCACGAGT

-Idefix

143

CGTTTACGGTCTTGTGGAATCTTTCTATATCAGATATACCATTTTTGCTAGTCGTTATA
GAAATTTGTACACCTTCAGATCTTAACCAATTTTGTAAAGGCTATACACATAAAAGCTGA
GTCTTTGTCTGCTTTAATTTCTTGCAGTTTTTCCATATCATTGAATATTTTAGTAATGG
CTCTTTTGTCTTCTAGCCAATCTCTACTTTTTACTTCAACTAGTGATGCAAATTCGAA
TAGATATCAATGCAAGATAAGAAGATCTGGTTTCCTGTGAGATAAAAAATCTATCACGTA
TTTTTCTCTTGTGTTAAATATTTCTGGTGTAGTCTCATATGTCAATTTTCGTGTTTCGAT
GTTCTGTTTTTGCTAGATTACAAATTTTACATTCATTGATAATGTTTTGAATAAGTTTTT
TGACTATCAGGATAGTAATATTCTTTCTTTAAATAAATTGATTGTTTTTCTATACCCGG
ATGTAAAAGTTTCTTATGTTTCTTTAAGATTAAATCTTTGAATTTGTCATACGTTAGTA
TATCAATTAACCTTTGTGGTACATCTCAAGAGTTTTGTGAAATTGTTAGGGCTTATAAAT
TCAGTAAACGCTCTCTGGAAGATCAGAAAATCTTCGTCATTAGGGAAATAAATGCACT
CTTTTTGGTACACACATATTCTTAAATGAGATTTTTGGCGAGTTCAGATGTCATTTCTT
TATAGACTATTTTAATCTTTAATTTTTGAAAGAAATGATGTACACTTGTGTATCGTTA
TCGCCTTTTTCTATCTCTATTTGTCGAGAGAAGTAATTTATTGGTCTTTCTGTTATGGA
TATGTAATTTAGATTATCTTCAATGGCACTGTGTATAGTTGCGTCTGCGCTGTTTGCGA
CCTCGCCAACCATAACTTCTTCTATTTTTCGTGCGGAAAGAGCATCCGCGACATGGTTTT
TCTTTGCCTGGAAGATATTGATTTTTATAATCGAATTCATTAAGTTTTTATTTTCCATCT
TTGCAATTTTCATGTTTGGTTCTTTAATGTTGTTGAGCCATACCAGTGGCTTGTGATCAC
TTAATATTTTCAAAAGGTCTGCCGAATAAGTATGACCTGAAATATTTTGTAGCCCAAAT
ATAGCTAAACAATTCCTTTTCAATTGTAGCATAGTTGATTTTCATGTTTCGTTTAGCGTTCT
ACTGGCATAACAACTGGCTTGTGATTTTGGGATAACACTGCACCAATAGCTACGTTGC
TAGCATCAGTTGTCAAAGAAAAAGTTTTGAAAAATCAGGATAGATTAAATATCGGGTCT
GAAGTTATCAAACTTTTAAATTTTTCAAATGATTCGATGTATTCTTTACATTTGTTGTC
TATTATAGCACCTTTCTTTAATTTGAGGGTCATGGGTTTAACTATTTTGACAAAAGTTAG
GAATAAACTTGCAGATAGAATCCACATAATCCCAAAATGATTTTATTTGCTTAGGTGTC
TTGGGTAATGGAATTTGTAATTGCTTTAGTTTATTTGGATTGGTTTGTATCCCATT
TGTTGTGACGATGTGCTTAGGAATTCAGTTTCTTTTTTTCATGAATTCACATTTATCTA
GTTGCAACTTTAAATAGCGTCTCTCAGTTTTTCAAAGACTTTCTTTAAGGATAAAATG
TGTTCTTCCAATGAAGTGAATAAACAATAATATCGTCTAAATAGACTAAACAGTCTTT
GTAGATTAAATCTTCCAGAAGATTATTCATGCATCTCTGAAAAGTAGCTGGAGCGTTTT
TTAAACCAAAGGCATACGAGTATATTCATAATGCCCATGCTTAGTTGAAAAGCTGTT
TTTGCAATAGAAATTTTCATCCATTTGGATTTGGTGAAAACCTTGCTTAGATCTATAGT
TGTAAGTATTGGCATCTACCTAGTTTGTCCAATATCTCATCCATTCGGGGAATGGGAA
ATTTGTGCTTAACAGTTATCTCATTTAGATTCTTTAATCGACTACCAACCTGAATTTT
TGTTTCCAGAGGCATCTGCCTTCTTGGGGACCACCCAAATAGGAGAACAATAAGGGGA
CTTCGATTTGCGAACAATCCCTTGTTCTATCATTCTTTAATTTGTTTGTAACTTCTT
GGTCAACGCTTTGAGGGTACTTGTAGGGTTTACGGTATACTGGGTCTTCGTGTTGAGTC
TGGATGACATGTTTAAAGTACTGGTGAAGGTCAAATTTTCGCCCTCTTTGTACTGAAT
GTCTCTATATTCGTATAGGACCTTCTTTAAACATTCAACTTCCTCTGCATTTAAGTGTT
CGAGTCTATACTCGTTACATTTCGCGTAACCTCATTGTTAATCGCGAAGTTGACTATATCG
TTGTCAACGGACTTGGGATCGAGATGTGTTACATCATTGTCCACTGTGTCTTTAACAAC
ATTTGAAATGAGGCATTTGGGTAATGCGGTCTCTTTCTTCTGCTGTTGATGCTTTGGTT
TAAGGCACTTAGGTGCGGTCTTAACCTTTTGCATTTTGGTTTGGTTCTGTCCACTAGA
GCGGAATCATTCTTTTCTTGTGGGTCAAGGCATTTGGATGCGGTCTCACCTTCTTTTCT
TTCATAGAGAACTTAAATACTTTTGGAGCTAGTGTTACTGTTTTCGTTAGCATAATCTA
TTTTAGCTTTTGTATCTTCTAAATACTTTTCGACCTAACAAGAAATCATAAGTTATCAGAA
AATTTGTGAATGTAGAATTTTGGAGCAGACGGACATACTGAAGTAGGTTTATTATTAT
ACTTTTCTTTAGTGTTATTGGTCCTTTAATAGTGATATCTATTAGGTTTTCCTCCTTTT
CTTCTAAGTTTTCAAATTTTCTCTTATGATGTTAATTGATGATCCTGAATCGATCATT
CCCTTATAATTTTTATTATGATGAATATTTTTACATAGGGACTTGATGTCTTCTCTCT
ATCATTTGAGTGTCGAGGCAGCTTGATGAAAATTTACATCCATCCTGCTTTGGTCACT
CTGACTCTCGCGTTGTCTTTTACTGGCCCTTGTCTGTTAATATTATTCGATGTGGAAG
GTTGGAAGTTAAGGGTGTTATTAAGTGGGTGTTCTGCTGAGCTTGTTGTTAAGTCTCGT
CTAAATGCATAATAATTCCTCTTTGATTATTTCGGTATGAAACGCGGTGAATATTGGTT
TTGATTTGTATAGTTGAGTCAATTCGAGGTGGTTGTGTATTTATGATTCTGATTCTG
TGCTAGGATAATAATTACTGTAGTTGTGATTTCTATTATTGTAATATTTTGTATTATTA
TTTTGATTATAGTTTCCCCTGTTTCTACGTTTATTTCATTTCTGTTCTATTGGAACGGTG

ATTTTCTGGAATAGCATCATAAAGGCCTAATTCATTGATGCTTGTTTTAATTTATAAA
 TTGTATCAATATCTTTACGTGCCAAAGACATATAAATTCATCCGGCAGTTTTCGATCT
 ATTACTGATTTAACAGTTTGTGTTAACATTTCCGGTATAATTCGATTTATCAACATGGTC
 ACTTTCTAATTCTAACTTATCGAACATAATCCAGTATTGTGTTCTAATTTTTCTATAAA
 ACTTACAGATACTTCTGGGTATGATATGTCTTCTAATTGTCTGATGAGAGTTACATAA
 GGTCTATGGTCTTTGAATGCCATTGCCAGTGCTTCTTCATGTCCAGCCATGTGTTAGC
 TCTTTCCTGGGTACGACTTCTAAGGCTGAATCCACTAATAACCGCTTCACTGCTCCAT
 ATATGACGTGTGCCGTCTTGGCTTCTGTGTTGGGTATAAACTGAGCAGTTGTTCCACT
 TGGTTGACGAATCCAGATAAATTTGCCAGGTGCGCCATCGTAGTAGTTAATTTCTTTGAT
 TTGGTTTAGCAGTTGGTTTAGGTGTGTTTCTGAGAGTTGTGGGACTGCCATGATGTCTT
 TTAGTTTTCTGGAGCTTATTTTGGTTTTAAGTTTTCTTGATTTTAAGTTTTCTTTAAAC
 TTGGTTTTAAGTTTCTTGATTTTAAGTTTTCTTTAACTTGGTTTTAAGTTTCTTGATT
 TTAAGTTTTCTTTAACTTGGTTTTAAGTTTTCGGAATTAATAAATTTAAATTAATTTTT
 GTTTCGACCGCACGGGTTTTTATTTTGAATACTACTTTCACGTAGATTCTTTTGC GGA
 TCTCTAAAATATTTTTGAATCACTATTTTAATTCGCTGATCTTGGATAGCTAGTCGTAT
 AGCATATCCTTCAGTTGATCAGTACCGTACTAAAAGGACTCTTTTATTTTATTGGATCC
 TACCGGCTGCGCCAATTAATACTTGCAATATTTATAAGTGAAAAAATAAATTTTTTA
 TTCTACTAAATGATTACAATCTTTTTATTTGAGAAGAACACCGACTTTTACAATTAGG
 TCTCAAACTGAACTCTGATAATTATTCTGATTTGATTGACGTTCAAGCCTCGGTATCGA
 GCTTTTCTAATATGGACTTTAGACTTTAGGATTCTGATCGTAAAGAGAAAAAGCTTTCG
 GTTCTCTTGATTCAAGTGCAATTTGCTCTTGGATACAAGTGCTCTTAGATTCTTATAA
 TTTGTTTATTGAAATCGATACTTTTGTGTTTTTCGGGATTGCGGCTCCGAGCTTTGAACGT
 GGGAAATGTGACGATGGGGTGAGGCTTAAGCAAGGAATTGTTTAAATTAGGGGACGGATG
 TTTAGTCTACCAAGTGGGTGACTTATCTAGAATTGGGGTTTTTCCATTCTCGAGGATCA
 TATAACAATCTTATTCTTATCTGGTATCTGTTGGGTACATCCGGAGTAGTGATGTTGT
 ATTGGGGTGTATATGTTGAAGTGATTAGTATGTAGGCATATGCTTAAGTCTTAGGGA
 CTTATGGATATGTCAC

2L38, -CG12717, -ade5

TATATAGTCGAACCAAAAAACAAAAAGTGTTCTACTTATACGAGAGCTAGGGCAGGACT
 TCTGTTTGCTCTGTAGTTGTATTAGCATTATGTATTAAAGCATTATCAGTATTTGATTT
 CAAACGTTGGGATACCAAATACGGTATAAACTAATTTATTAGCTTCTGTTAAAAATTGT
 ATAAATCGCGCTTTGAAAATGGTAATGTATTTGGGGATTTTAAATCCCGTACCAGTTTCG
 TGAAATAAAAAAAAAAGCTGATAGTTGGGAAGCCGGTCAAACTTGCATCCTTAAATTA
 AAAGACATAAAGATCGTGGCGCGTGATGCCTTTGTGATCGTTAACTTCGTGTTTCGAAA
 CAGGAAATTTGCCTTGACATAACAGACTAGTTGTTGCTGCTGTATCTGGGCCAGAGACA
 TGATTTACGAAATTTTAAATAGATGTCAATGGAAGTCGACGATGTCGTTCCCTGGACACCG
 TCAACTGACCAAGGATGCGTAACGCGCCTTCCCCGAAGCTACCACTACCGTGACCGCC
 TTCCCGGCCGGCAGGCATACTTCTCCGATCGGTGTCGGATACGGAGTGACAAATAGCT
 GACAGACGCCCGCTTATATAGCTACTGGCTGGGCTTAGTGTGCCCGCCGTCTGTGTGG
 CATTGCGCTATATTAGTGTGGCCAGCCTGCTTTGTCCAGTGTGACCCACCGCCGATCT
 TATTGGCGCACTCGCTTAAAAAATAATAGGAAATCCTTTTGTAAATAGTTACCTAATT
 TGATACTGAGAATTTACTGCACGACCGGAGTAGCAAATAGGGCTTTTCAATGTTTTTGT
 TTAAAGAGGAGCTCAATGCTGACGACTAAAAATCGGACAGGTGGACCCGATGAGGGACG
 CACAATAAATTTGATTAATGTCTGTTGAAATTAATACGAGAATACATTTTGTGTTTTT
 GAAATTTCCATACAAAGAACCTTGCAACAAACGGCTGAACTTTAACTTCGTAAATACCA
 CATGTAAGAATACTAATCTTTGTAAATGAAAAGCAAAGAAGCTTAGCTTAGCTTAGACTCG
 CTTAAATGCCTATATGTGGGGTCAGTTCAAATATTTAATTTAATATTCGCAACATG
 CAATTTTAAAAGTGACTATCGATACATAATTTTAAAAACACCCACCAAATCCATATC
 AATTGGAAACATATATGTATCGATATGAGCAGCAGAACTCCGATGGCGATAGAACTCG
 CACAGATGATTTTTGGATACAGCACTTGAACGCCGTTTACCCCAATGATAAAAAATAGA
 GAGATATTAATAGCGCATGAATCACACAGGTTGCAAAATGAAAGGAATATAGTGATCA
 ACGCCACCGTGTATGGTGACTCTTCAACCATCCAGTTAACTTTGCACATCAAGCATCCC
 GAGAATTTTCAATCGAAACGAGAACTTCATCTCAGTCGGATTAGCCCATACGTACAT
 ATGTTGGTTGTTGAATCCGGCTCCAAATGGATCGCAAAGAAGCTGTTTTGAATGACCA
 AGCGGTACGTGAAGAGCTCTCTCAACGGTTGTCAAACCATCAGAGACTTTGGTGGCCA

ACAATGGGTGCCAGCCAGCTGCAAGTGCAACCAAAGGCAAGAAGCACAAGAAAAGTCCT
 TGCTCTCCGGGAGACACCAATGCTGGCCAAGTGATATCGTCGGTGGATGAAGACACCAA
 TGCTGGCCAAGTGATATCGTCGGCGGATCCAGACGCCAATACTAATAACCAAGAACAAGA
 AACAGAAGAAGTGCCATTTTGC AAAAGGATAACAATGGCTATCAAGCTGCTGGTTCTCAA
 GACAACCAAGTTTCATTTGGCAATAAACAGGGAAGTCACAACCTCTAGGCTATAACCTCC
 AATTCTTTAATGACGAGTAATGGTCTTCAGGGCATGCAACCAGCTGGAGATGCTCCCGC
 AATCTCCCAGTGAGTCCAGTCGAGGAACCATGCCGCACATGCCTCTGTTAGCAATAAAT
 CGGAAGAGACAGCTGGTGGATCTCAATCGCGGGTGCAATTGAATGTTGCATCCATACCC
 TCCCCAGCGATGAAAGCAACTAGTAACGCAGCTATACCCACACCGCAGAACGCGCAGA
 GCGGAGTCACCTGCGCCGGAATCGCAAGTGGATACTAAGCAGGGACGTTAATGAAGATG
 CCGTGGTGCTGGTGAGCAGCGGCTATGAGGAGACCACCGCCGCCGACGATGGCCAAACA
 CAGAGAAGACTCTCACCCGACGAGAAAACCTTTTTCACATATTCGCCAACCGGTACGGGT
 GGCTGGTATATACTCTTCTTTACAATTCTCCTCCTCGCTATCCTCATCCATCACTAGT
 TGCCCATTTAATGTGGGAAACTTAATGACCGGCAAAATTTTGCCTGCTGATTGCTTTTC
 ATTCATTAGCTTCTGGATGAGATTAGCGATATCCTCACGCTTCTTAGTTACTGTGAGCA
 GTAATTTTCTTGTGGAAAAATGTGCTAAAGATGTGCGTCTTTAGCCAGCTGATGTTATT
 CTTTAGCCATCGCAGATGAAAAATCAATGAATATATCATTAGATACGATCCTTTACTGA
 GACATGGTGACATCGGCCAACAAAGACGATAAAGAGGCTACCTACTAGAGGATGAGCATG
 AAAGGCAGGATGAGCTGGATGTGAGGAGTTGGAAAAATGGCCAAAAATCAAGAAAAATTC
 AAAATTTAGAAAGTCACCAAATGAATAGAAAATAAAAAAAAAATCGAGCCTTAACGTAG
 TATAGACGCTCAGTTGCTTTGGAGCGATTCTTTGATGCAGACGAACGTCGAAGTTTCT
 GATTGTAATGATTGTGGGTTTTGAATAAAAGCGACTACGTTTTTATTTTCGATCTATTTTC
 GTATGCAGTTATTTTGATTTGTTTTCTTCTTTTATTACGCGTATTAGAAAAATAAAT
 AAAGCTGTTAGTTAAATATGTGGATTTCGTGGATTTCGCATTTCGAGATTTCGCG

Doc

GACGTGTTTCTTTCAAGCTACGAATAGCAAGTTCTAAAACTACAACAGTATAGTGAAA
 GTTAAACACAAAGTGTAAGTGAGTTTGCACAACCTAACAAATTATTGACTATAGTAATT
 ATTTACTAAAAATAAATAATTATTCCATATTGTTCTGGTAATTGTTATATGTGGACTTAG
 AACAAATGAATCAAAACGACATACGTTCTCAGCGACAATGTGAACAAGACGAGCGCCGGC
 TCTCTTTACAACGCAACAATGCATACTTTTCTTTCTCGTCTCACCGCAAATCGGTGATCGA
 GCACCTCACCTTCAACTAAGTCAAACTTTTGGCCTCAGAGAACGTGAGACCGCGTTC
 TTGCTCTCCCTCTCTGCTGCTTCGGCTCACAAGTCTGGAGCGAAGAGACCGCTCTC
 CTACCCCGCTCCTCTCGCAGCGCCACACGACCGTCCCGGTAAGTGTAACTGTAACACTGCAATA
 ACGAGTGACAGTGACCTCACTGGCAACTGCCACTGCTACCACAACATCAACTTCGTCAGC
 GGCCCAACTAATTATCGCTGTGCGAGCTGTAATAAATTCAGCAGCACTGACCGCTTTGCA
 ACAACAATAATGCAGTCAAGGCAAAAGCAAAAGCAGAAAGTCGATTTCGACTGTG
 CAGACTGGCATTGGATCGCTACATCAAAATCAAGAGAAAGCTCAGCCCTCAATACAATAA
 GGCAGGTAATCAACCCAAAATCAATCGAACCAACAACGGCAATGAAAACTCTGCAGTAA
 ATAATTCAAACCGATATGCTATCTTGGCTGATTCTGCGACCGAACAACCCAACGAAAAA
 ACTTTAGGGGAACCAAAAAAGACAGGCCTCCACCAATTTTCATACGAGAACAAGTAC
 AAATGCACTTGTAATAAACTCGTTGCTTTGATTGGTGACAGCAAAATCCACATTATCC
 CACTTAAAAAAGGAAATATTCATGAAATAAACTACAGATCCAAACAGAAGCAGACCAC
 CGTATAGTGACTAAATACCTAAATGATGCTGGTAAAAACTACTACACATACCAATTTAA
 AAGTTGCAAGGGCTACAGGTAGTACTTAAGGGCATTGAAGCAACAGTGACACCAGCTG
 AGATAATTGAGGCTCTGAAGGCCAAAAACTTTTCTGCAAGACAGCTATTAATATTTTA
 AACAAAGACAAAGTTCCGCGAGCCACTATTCAAAATAGAACTCGAACCCAGAGCTCCAGGC
 ACTAAAAGAAAAACGAAGTGACCCCAATATACAATTTACAGTACTTGCTACATCGGAGGA
 TCACCGTGAGGAGCGGCACAAACGTACCAATCCAGTTCAATGTACTAATTGCCAAGAA
 TACGGCCACACCAAGGCATACTGCACCCTTAAGTCCGTATGTGTTGTCTGTAGCGAACC
 TCATACTACCGCAAACCTGCCCCAAAAACAAGGACGATAAGTCTGCGAAGAAATGCAGTA
 ACTGCGGGGAAAAACATACTGCAAACTACAGAGGCTGTGTGGTGTACAAAAGAAATTGAAG
 AGCCGCCTAAACAAACGTATTGCCACAGCACATACATACAACAAAGTCAATTTCTACTC
 TCCGCAACCGATTTTTCAACCACCCCTAACTGTCCCAAGCACTACTCCAACAATTTCTT
 TCGCTAGCGCCCTAAAATCCGAGCTAGAAGTGCCCGCCCAACGACAAGAAGTGTCTCAT
 TCCGAACATACACCGACAAACATCCAACAACACAACAAAGTGGCATCGAAGCTATGAT

ade5, CG12717

147

GCAAAATACATTACTGCAGCCAGAACAGATAAGACACTATTTAGTAAAGTATATATATT
 AGTAGTAATAGTAAATTTGTGAATGTAAAACAAAAGACAGTTTCTAGTTCCTTCACATT
 TAAGGAATTCGACTTGGAAAACCTTTCTGACAGCTTACAACGTTGTGATTTAGATGCTTT
 TTTCGCACCTTGGCAACCGATGCAAATTGCGAAATGGAGCCCTTTGGCCCGCGGAAAGG
 AAAACGTGTGTCTGTTCAGTTTTTTTTTTTTTATAAAGACGCTCTTGGCCAAAAACAAAATT
 CCCTACATTCAGATTTCAAATCGGACTATTGATAAATATTTTAAAAACAACAACCAAAA
 CAACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA
 TGTATCGATATGAGCAGCAGAACTCCGATGGCGATAGCAACTCGCAGATGATTTTTGGA
 TACAGCACTTGGGACGCCCCTTTACCCCAACAGTTAATAGATGCATAAATTAATATACAA
 GTGACATTTGTATTACTTCTAGTATAATATATATCTTACAGTAACTTGTGGGCAATCAA
 AGCCATAAAAAAGAGCTATTGCCTTGGCATGACATGAAATGAAACGAATTTACGTATAGC
 ATCAGCACCATTGCCCTGGCGAAAATCTTGACTTGAATTATAGATTTCAAGGGCTTTGA
 GTTCACAAACCGGCCAGATCGTGTTCTTTGACACCATCGCCGGCGGTGATGCGATCCTT
 GCCAAGGAGCAGGCAAAGTCCTAGCTGCTCTGGCAGATCGTACACCGCTTGGTTTTGC
 CTTCGATGATGACCTTGACGAGCTTATAGTCTCAACTGTGGAATCCAAGTTGCAGTT
 GCATTCGTTACGTTTGGTGTTTAGTTAGTATTTAGAATCGATGCTATATGATATATGAT
 TTAAGGAATGGAAGACAACCTTGTTTACAAAATCCAATATTATCTTGATTTAAACAATG
 GAAAGCCGGCCTTCTTGTGTGACGACAGCTCAATAAACAGCATTAAAAAATGATAG
 TACTTTATTCTTGATTATCAGAACCAGGAAAATATTCGTTGTTTTTTTAAGTAACCCG
 AATAACAACCTGCTTAAATAAATCCAAGGCTGGCGGGAAAACATAAAACAACCTGTCTTGCT
 TTCCACCACATTAAGCTTGTCTTTTATTGATAAAAGATCGGATAAGCTTTAAGGTAGCT
 TTGGTAATATTTACGTTCTAGGCACAACAATTGTTTATTTGGACTTGTGGGCTTTATGA
 AGTTTATAGTTTATGGTTTCGGGAAACCTTGGCAATGGATAAACTTACTGGACGCGGT
 CGTGGTGGTGGACATCTCGTTAATCGCACGTTGTGGTTGCTTCACTTTAAGAATCGAA
 ATTGAAATCGATTGGCGGCTGTCTGTTTCACGTGTTGCGGGCTTTGCTTTATGCTCGAT
 ATTCGCTGGTGCCCTTCCCCCTCGCTCTCTATCTTTTTCGCACTCTCCGCTTGGCTCTAT
 TATTACAATCCGCAAACTCCGACGTTTCGCTCGCATCAAAGAATCGCTCCAAAGCAACTG
 AGCGTCTATAGAATCGCCGGCGGACTTTATGGCTCGATTTTTTTTTATCGCGGAACCCC
 AACACCAATGCAAAGAGATCCATGTGTTTGGCTCGCGTGTGTATGTTTGGCTGTGTGT
 GTGTATGCTTAGAGTCGAGTATAATGCTGATTGGTAATACCCTGCAATTTCCCAAATGG
 CTGATGAACACGGCGCGGGGCCATTTATGTACTTATGTACATACATATGTACAGTTTTG
 GCAATCTATAAAGCTTTGCTTATGCGTGACAAATTTTATTGTAGCATTTTATTTACAAA
 TTTTGTCTTCTAAAAATATTTTCTATTCTTTGGTGACTTTCTAAATTTTGAATTTTTTTT
 TATTTTGGCCATTCTTCAACTCCTGACATCCAGCTCATCTGCCTATCATGCTCATCC
 TCTAGTAGGTAGCCTCTTTATCGTCTTGTGGCCGATGTCACTATGTCTCAGTAAAGGA
 TCGTACCTGAATGATATATTCATTGATTTTCATCTGCGCTGGCTAAAGAATAACATCAG
 CTGGCTTAAAGACGCACATCTTTAGCACATTTTTCCACAAGAGATTGACCACACGAACGA
 ATCCCACAAATATGAAGCTGACGGCAGCCCAGAAGCGTCACGAAAGGGTTGAGGAGTGG
 ACACGAAACGTGAATATATTCGACAAGGACTTCATCATAATACCATTCAACGAGCAGTC
 CCATTGGATACTGGCCATCATATGTTATCCAAACCTTGAATCCCGTCGGTACACAATA
 ATAATGTACAGACGACACTCTGCGATGATATCCAATCAAACAGCATTGAGTCTTATTT
 TCGATTCTGTTGGCGGTACCTTACGACATCGAGCGATTGCCATATTGCGGGATTGTCTC
 ACCTGTACAAGGCAAAGTATCCGAATGCGCTGGCGCATATCTTCAACAAGGATAATATG
 CCCGCTCATAAAGTAGAGGTGCCGAGCAGCAGAATCTCACAGATTGCGGCCCTCTATCT
 GCTGCAATATGTGGAGCAATCTTTTACAAAACCTTTCATCGACTATACACTGCCCATTG
 GGGAGCTGAGCAATTTGGTTTGACATGCTCACAGTAACCTAAGAACCCTAAGGATATCGCT
 AATCTCACCCAGAAGCTAATGAATGAAAGCAATCAGCAGCGCAAAATTTTGCCGGTCAT
 TAAGTTTCCCACATTAATAGCCAACCTAGTGATGGATGAGGATAGCGAGAAGGAGAATT
 TTAAAGAGCATTTACCAGGCCACCGGTACCGATTGGCGAAAATGTGAAAAGGGTTTGAT
 TCTCGTGGTTGAGAGTCTCTCTGTTTGGCCATCGTCGGCGGCGGTGGTCTCCTCATC
 GCCGCTGCTCACCAGCACCACGGCATCTTCATTAACGTCCCTGCTTAGTATCCACTTGC
 GATTCCGGCGCAGGTGACTCCGCTCTGCGCTTCTGCTGGTGTGGGTATAGCTGCGTTA
 CTAGTTGCTTTCATCGCTGGCGAGGATATGGATGCAACATTCGATTGCACCCGCGATTG
 AGATCCAGCAGCTGTCTTTTCCGATTTATTGCTAACAGAGGCATGTGCGGCATGGTTCC
 TCGACTGGATCCACTGGGAGATTGCGGGAGCTGGTTGCATGCCCTGAAGACCATTACTC
 GTCTTTGAAGAATTGGAGGTTATAGCCGTAGAGTTGTGACTTTCTGTATTATTGCCAGA
 TGAACCTTGGTTGTCTTGAGAACGAGCAGCTTGATAGCCATTGTTATCTTTCCAGAAT

GGCACTTCTTCTGTTTCTTGGCCTTGGTATTAGTATAGGCGTCTGCATCCGCCGACGAT
 ATCACTTGGCCAGCATTGGTGTCTCCCGGAGAGCAAGGACTTTTCTTGTGCTTCTTGCC
 TTTGGTTGCTCTTGCAGCTGGCTGGCGACCATTGTTGGCCACCAAAGCTTTTGATGGTT
 TGACAACCGTTGGAGAGAGCTCTTCAAGTACCATTGGTCCTTCAAAACAGCTTCTCTG
 CAATCCATTTTGGAGCCAGATTCAACAGTACAACCAACATATGTACGTTATCGGCTAAT
 CCGACTGAGGTGAAGTTTCTCGTTTCGATTGCAAATTCTCGGGCTGCTTGATGTGCAAA
 GTTAAC TGATGGTTGAAGAGTCACTTACACGGTGACTTTGATCACAATATTCCATT
 ATTTTGCAACCTGTGTGATTTATGCGCTTTTAAATATCTCTCTATTTTATCATTGGGG
 TAAACGACGTTCCAAGTGCTGTATTCAAAAATCATCTGTGCGAGTTGCTATCGCCATCG
 GAGTTCTGCTGCTCATATCGATACATATATGTTTCCAATTGATATGTGTGCGCGGATCGA
 ATATTGTTATCGATAGGCTAGTTAGTATTTTGTAGAAGTCCGAATGTGGAAGGATTTGT
 AAGCCTATATGTGTCTGGGCACATTGTTTTTCGCCATTGTTAATTGCCGGGAAAATTTA
 GCTTTAGCTAAATTAGCACACTGCGATGAGCTAATAAGTTAAGTTAGTTGCAATTGTGA
 AACATT

aurora

GAATTCCTCCAGAATAAAACGTGTTCTACTACCACGGATTAGTCTGCCCTTTCTTTTCGG
 GAACCAATGCGTGGGGTAGCCGTTTAAAGGCAACTCCATGTGACGCACGACGACGCTT
 TTATTCGGAGTCCTAGGGCGACTGCAGGGGCAACTTGCGCTGATATGACGGTTTAGACG
 GACAGCTAGAGAGTTGCCGGAGCTGGAGTGACGGTTTAGACGGCCAGCGAGGATGATTT
 GTGTGAGCGCAGCCAGCGACAATGGCGTCGCAGTCAGCGACATAGAGGGACGCAGCCAG
 CGTCGAACGCCGCTACGAACGGGTGCGAGCCAGCGACAAGGAGACGCAAGAAGCGTCAT
 TTGTGGAGACCGCAGCCAGCGGTAAGAAGGCGCAAGGAGCGACAAGCATTCGTGGCGGC
 AGCCAGCGGCACGAGGCGTCAGAGACGCCATTTTGGACGCGCAGAGGCTCCGCCATTTT
 GGAGCTGGGAAAAGATGGAGCATTTCCCCCAGGAAGAGTGCCCGGCTGAACGGAGGGGAAG
 CCACCCCTATAACAACAGTGAGTCAGCAGCGAGCCAGTAGTGGAGCAGGAACCTCGGACG
 CGGGTGAACATCACGGCGGCGTCGATTCTTGGCCGGCCACTACGGTGACTACAGTAGC
 TTCCCAACCTAGAACTACTGCTGTACAGCTGCGAGTTTCACTACAGGAGTGAACAGC
 CCTTCGCGTTTGGAACTCATGGAGAGGATCGCAGCGTTGGAGAGGGAGCTGGAGAAGGCT
 AGATCCCTGGAAAGTGTGAGCACCGCCAATTGCGCGCAGTTGGCCCAAGCGCAGTTGGC
 GCCAACAGTGAGCGTCGGGGCGGCGCCATTTTGGATCGGCCAGCCAATACCCACATC
 TAACGGAGAGGCCTTACATAACGGGGTGGGTGCGTGCATACAAACGGTGACGGTGCGA
 GCGGTGCGGCATGCATGCTGCCGAATCTTCGAGTGGACCTCCATGCTAACGACTATC
 AACTATTTTGTGGAGCCACTGTGTGCAACAGGCAC TGCGCAGCCAGCGCATGGACTCGT
 GCTACCGGGCGTGAGCATCCACAATGCGGCAACAGCATCGCCACTTGTGCGATCCTACG
 CCGCGACGACGCCGAGTGGAATCCAGGGAGCATATGGGCCAAGGAAGCTTCCGGACTTG
 CCTATATTTGGAGGAGCGCCGAGGAGTGGCCGATCTTCAGCTGTGCGTGTGCTGGAGAC
 GACCCGAGCGGTACAAC TGACGACGACCTGGAGAACAACAGAGGTTGTGGAAGCGCTGA
 AGGATGAAGCGCGCAGGCGAGTGAAGGCGCTATTGATTCATCCAGGGAATGTCAGCGCC
 GTGATGGAGCAGCTGCGCTTTAGGTTTCGGCCGACCGGAGCAGCTTATACGCAGCCAGCT
 CAACAACGTGCGAGAGGTGCAGCCAATTTTCGGAGCACAATTTGGCGAAGATCATTCCCT
 TCGCAACCCGAGTGAGTAACCTCGCGGCCTTCTTGCAGTCAGCGAAGGCCAAGCAGCAC
 CTGGGAAAACCAACCTTCATGGAGGAGCTTGTGGCCAAGCTGCCAACGAGCAAGCGAGT
 GGACTGGGCCAGGCATGCTGCAACGATTGCGCCCTTTCCCACTGTAGTCCCATTACGCG
 CGTGCTACAGGAGTACGCAACGTTGGTGTGCGTCGACTTCTACATGCAAGCGTCGACC
 ATAATGAATGCGATCAACAGGATGATCGGCATGGAGGTTGTCCCATCTGTGGAGGACAG
 CATGGAATATTGAACTGCAGAGAATTTATTGGAGCTTCGCCACAGGAAAGGTGGAGCAA
 TTTGAAGAGGCATCGGCTCTGCTTCAATTGCTGCGAAGCGGGCACACGGCTAGATCCT
 GCTATACGCAAGGTGAGTGCCAGATTAATGGATGCCGAAGGGAGCATCACCGTCTGCTA
 CATGGTGCAGGACGAGGAGCGAAGGCCGCTGCAGCGAGGTGGCTTCAGACGCCACGAAGG
 GAACAGCAGCCAGCAGTTTCCAGACGCAGCCCGGCCAGGAGGCTTCGCTACGAGATG
 GTCACAGGGACCAGGAGAGGAACCGGCAGCCAGCCGTTCCAGCAACAGCCCGGAGAGG
 AGCTCCGCGTGAAGCGGGAGCGCCCTTGCAAGGAATTTGAGCTGCGTTGACGCCGAAG
 GAGGCCGTCTACTGTTCCGTATACTGCCGTTACGCTGAACGGAGCGGGGCGAAAGGTA
 GATACGTTTGCCTCTTAGATGAGGATCCTCCGTCACGATGATCGATGACGAGCTACG
 AAGGGATCTTGGAGTGCAAGGAGAGCGTCGGCAGCTAAATATCCAATGGTTTGGAGGTA

AGGCAACCAGAGAGCCTACCAATGTGGTGAAGTCTGAAGATAAGTGGAGTTGGAAAGCCC
 ACTCGCCATGTATTGAAAAACGTTTATGCCGTTTCGAGTTTGAGTTTGCCGATGCAGAC
 ATTGAGCCGACGAGATGCCCAGGGCGTGACAGGGATGCGCGTCTGCCGATGAAGCCTT
 ACAGCAACGTGGTGCCGAAGCTGCTCATCGGCCTGGATCACGGACATCTGGGGTTGCCA
 CTAGGACAAGGCGGTTTCGCTCGAGAGGGACCGTATGCGGCCGCAACCGAGCTGGGCTG
 GGTGTGTTTGGGCCTGTAAGTGGGCAACCGACAACGCCACCAGTCCAAGTAGGCT
 ACTTGCCGTGTCACTGGATGACGCGATGGAAGATGGTGGTGGACTATTTTCGACATGG
 AGAAGTTTGGAGTGAAGACCGCGCCGGTTCGAGCCAGCGACGATGTTTCGGGCCCCAA
 AGGATACTCGAAGACACCACGGTGAAAGTGGGGCGTTCGTACAGACGGGATTACTCTG
 GAAGGACGACCACGTTGTAATGCCACCGAGCTATGAGATGGCGTACAGGAGGCTGGTCA
 ACGTCGAGAAGAAGATGAAGCGCAACAAGCCGTTGGCGCAGGAATACGATCGGATTATA
 AAGGATTACGTGTCTAAAGGATACGCGAGGAGGCTGCAGCCGAGGAGGTGCGGGTAAG
 GAGCGATCGCCTATGGTATTTGCCACATTTGGGTGTGAAAACCCAAACAAGCCCCGCA
 AGCTTGTGTTTGATGCTGCAGCCAAAGTTGGAGGAACCTCGCTAAACTCGGAGCTGGAC
 AAAGGGCCTCAGCACTATAAGCCTTTGCCAGCTGTGCTCTTTCATTTTCAGAGAGGGAGC
 CGTCGGAGTCTGCGGTGCATCAAGGAGATGTTCCACCAAGTGCTGATCCGACCCGACG
 AGAGCGAGGCTATTCAGCTATCTACCCGAGTGAAGGAGATACACAAGGATGCTGGATTTC
 GAATTATGCCAGTATTTTCATCAGCTACCCACCGTGGAGACGGCTTTAGGACCTGGTTCG
 AGTCAAGAGCGTCGGATGAGGTGAGGCTGAAGAGAAGATCCTCGGAATGCGTTGGCAAG
 TAGCAACAGATGACTTCAGATTCAACGTGGAGTATCACCGAGTGCCAAGCAACTTCCTG
 AGTGGAGATCGAGTCCCTACGAAGAGGGAATATTTGAGCCTGGTGTGTAACGTTTGA
 TCCCTTGGGATTCTGTGCTGCCTCATGGTTACAGCGAAGCTGTGCTGCGAGAGATTT
 GGAGGCAGAAAGATCCAGTGGGACGAACCACTACCGGAGGAGTTAAGCAAAAGCGTTTGCG
 ATTTGGCGCAAAGAGATGGACGCCGTGAGACAGTTCGGATGTCCGCGCCGGGCTCTAAG
 ACGGTGCTGAGTTGGATCGGCAGCACCCACCGCCGGTATAAGCAGTTTGCTGGCAACCG
 AGTGGCGGAGATTTTGGAGTCGTGCAAGGTTTCCCAATGGAGATGGGTGCCCTACAGCCG
 ACAATGCGCGGTGATGTCGACGCGCGGCGCAGAAAGGAGTTCGACCTTAGCCAGGAATCA
 AGGTGGCTAAGAGGACCTGCATTTTTTGGAGCAGCCAGCAGCCAGCTGGCCGCGGCTGA
 GGAAGGAACTGAGCGTGTTCAGATGCCACTGATGAAGAAGAGGTGCCAGTGAGTTTG
 CATTAGTTGCGGCAGACGATTTTGTATTTCGTTTCAGAGATTCTCGAGCTTCAGTCGC
 CTGGTGAGGACCACAGCCTGGGTCTACGGTTTGCAGCGTGGTACCGCAAACAGCGAAA
 CGAGCTCGAGGAATACGGCCTTACTGCAGCAGAATGTAAGGCCGCGGAAAACCTGTTGG
 TCAGACAGGCACAATTGGAGTCGTTCCCGACGAGATGAGGTAGGCGGAAAACCTGGTCAG
 GACGTCGCTAGATCGAGCGACATTCGAGGGTTGGTGGCTACCTAGGCGAGGACGGGAT
 TCTGCGAGCTTACGGCAGAATTGATGCCGCACTGTGCATGCCGTACAGTGCAGGAGGC
 CCGTATTACTGTACACAGGCACAGTCTGACAGAGCTGATTGTGAGAGACTTCCACGCC
 AGCATGAAGCATCAAAATGTGGATGCTACGATTGCGGAGATCCGTACAAAAGTTCTGGGT
 CACAAAAGATGAGGCGTGTGATGCGGAGAGTTATCTCATCTGCAACGAGTGCAAGTTGC
 AGCGAGCGCGCCGATGTGCGCGATAATGGGGCCACAACCGGAAGACAGACTGGATGCG
 GGTGGATGGCCATTCAAATACACAGGACTGGACTACTTTGGGCCACTGCTGGTGACTGT
 GTCCCGTCACAAGGAGAAGCGTTGGGTGCGCCTGTTTACGTGTTTGACGACAAGGGCGA
 TTCACCTGGAGCTGGCGCATGACCTGTGACGGATTCCTGCATAATTGCGATCAGGAAC
 TTCGTCTGCCGAGAGGGCCAGTATAAAGACTGCGCAGCGATAACGGCAAGAACTTCGT
 GGGAGCTGACAGGGAAGCCAGGCGCTTTGGTGACGTATTCGAGATGGAGAAGCTTCAGA
 GTGAGTTGTCAAGCAGAAGCATTGAATGGGATTTTAATTATCCAGTGAACCCGCTGAG
 GGCGGAGTTTGGGAGCGCATGGTGCAGTGCCTCAAGAGAGTATTGCGTCATACCCTGAA
 GGAAATTGCGCCGAGGGACCATGTATTGGAGAGCTTCCTGATTGAGGCGGAGAATATTG
 TAAACTCGCGCCCGCTCACCCACTTGCTATGGATGCGGACCAGGAGGGTCACCCACC
 AGGGTGATATGGTCTTCGTCTGCGATCCCGCCTTGGCCGACGAGAGTGGCGCAAGGGC
 ATCCTGGAGGAGATCTACAGCGGAGCTGATGGAGTCTGTCAGACGCGCTAAGGTGCGCGT
 GAACGACAACGGCCTATCTAGGACAATGATGCGACCCGTCTCTAAACTTGCAGTTTGG
 ATTTGAGTGAAGCGGTTCTTCACGGGGTTGGGGATGTCGCGGATCGAATATTGTTATCG
 ATAGGCTAGTTAGTATTTTGGAGAAGTCCGAATGTGGAAGGATTTGTAAGCCCATATGT
 GTCTGGGCACATTGTTTTTCGCAATTGTAATTTGCCGGGAAAAATTAGCTTTTCATTGT
 CGTGTAAAGATTGGAGGACACACTGCGGTGAGCTAATAAGTTAAGTTAGATGCAATTGT
 GAAACATTGAATTCTTCCAGAATAAAACGTGTTCTACTACCACGGATTAGTCTGCCCTT

TCTTTCGGGAACCAATGCGTGTGGTAGCCATTTGAGGCAACTCCATGTGACGCACGACG
ACAA

-2L38, 2L35, -2L38, 2L35

TATGGATTTTGTGGGGTGTTTTTAAAAATATGTATCGATAGTCCGCTTGAAATTTGCA
TGTTGCGAATATTAAATTAAATAATATTTGAAC TGACCCACATATAGGCATTTAAGCG
AGTCTAAGCTAAGCTAAGTTCTTTGCTTTTCATTTACAAAGTTAGTATTCTTACATGTG
GTATTTACGAAGTTAAAAGTTTCAGCCGTTTGTGCAAGTTCTTTGTATGGGAATTTCAA
AAACAAAAATGTATTCTCGTTTTTAATTTCAACATTTTGCAAATACAAGTACAGACACCA
ATCAAATTATAGTTGTGCGTCCGTATCGGGTCCACCTGTCCGATTATTAGTCGTCAGC
ATTGAGCTCCTCTCTAAACAAAAACATTGTGAAACACTATTTGCCACTCAAAAAAGGATT
TCCTATTATTTTTTTAAGCGAGTGCGCCAATAAGATCGGTCCGTGGGTGTCACACTGGACA
AAGCAGGCGGGCCACACTAATATAGGCAAATGCCACGCAGACGGCGGGCACACTAAGCC
CAGCCAGGAGCTATATAAGGCGGGCGTCTGTAAGGTATTAGTCACTCCGTATCCGACAC
CGATCGGAGGAAGTATGCCTGCCGGCCGGGAAGGCGGTACGGTAGTGGTAGCTTCGGG
GAAGGCGCATTACGCACTTCCTGGTCAGTTGACGGTGTCCAGGACGACATCGTTCGACTT
CCATGACATCTATTAATAAATTCGTGAAATCATGTCTCTGGCCAGATACAGCAGCAAC
AACTAGTCTGGTATGTGCAAGTCAATTTTCTTGTTCGAAACACGAAGTTAACGATCAC
AAAGGCATCACGCGCCACGATCTTACTGTCTTTTAATTTAAGGATGCGAGTTTGTCCG
GCTTCCCAACTATCAGCTTTTTTTTTATTTACGAAACTGGAACGGGATTAAAAATCCCCA
AATACATTACCATTTTCAAAGCGCGATTATATACAACCGTATTTGGTATCCCAACGTTTG
AAACCAAATACTGATAATGCTATAATACATAATGGTAATACAAC TACAGAGCAAACAGA
AGTCAAACAAATTTTTAAATTTTTTCCATGTTATTTCTTAATATCTATCGATATCAGAG
AAAAGTAATAAAATTTTCGCGTTCGTTTTTACACTAGCTGAGAAAAC TCGAACAACAAA
CAACAAATTATAAGGATGATGATACCGATTAAAGGCAATCTATTCGTGCTCCTCCTCCT
CCAACAGATCGCTTGGCAATTCAGTGTGTTTGATCCAGATTCTTATGGCATTCTTTTCG
ATCTGACTGAATCGCTTTAAGTCTTCAGCATTCTCCACGCTTTTGCCGCTTTGATATC
AACTCTTGTGGGTTTACGTCTCAATGCTTTTTTCCGGTTTTTTTTTACAGTACGAGATAT
GTTTTTCCGCATGTGAATGTTGGCACCAGAATACGCAGGGGAAGAGAAGGCTTCTACAA
GGATAAAAAAGTTAAGCCCATATTAGGTATAATTCAAAATTATTCAAGTTTCTTGAAAT
CCGAGTTAATGGCACCAAGCTGATATTTTGGGCTTGACAAACATTAATTTAAAATACTA
TATAAATTTGAAATCTTCATCACCAGGAGTTCCTCTTTAGAAATAGCTCGATATTTCAAT
CGGCCAAACCGCGTGCATATGATTTATTTAAGTGCCCGCGTTTGTGCCAGTGTTCTT
GCAGCATGTGAATGTGTCATTACAAGGTGTGGGGTGGGCAAGGTGTTTGGCTGTCAAT
GGACTGCCCCCGGCATTAGTTTAAAGCCAGATAACGTTAGTTTGTAAAAATCGATCGTA
ATCCGTTTTATTTCACCTTACCCTCTGAATTAAGTTTTTATTAAACTGAACTGAAGAAG
AATTTTTTTGGAATATTTTTGTGCGTAGCTTAAATATAAATATTATGAGATATCGTTCC
TAATTATAATGTATAAAAAATGTACCGTGGAAGATGCGCATCCTGGGAAACATCTTTT
GTTAGGTGGCGTCGTAAATTTCTACATAGGTGGCAGAAATTATATGAAACTATTTTTATT
CCCCTATTGCTCGTGCCAACTCTTTTCGGCTAACAACTCGGTAGAGCATAGATATA
CTTTAAATGGTCCGAAACTACCTGTTACATACTTTTCAAAGAGTGTAGTATATACTTTC
ACTCTACGAGTAACGTGTATAACTACGTGTTTTAACCTTCTAACTTGCGTCGCCAGACA
TTAGCTATACAAACGATTCTACTATTTTTATTACTATTATAGTTAAAAATATTTATAATT
GACAACACACTTTGACAAGATCGCTAAATAATTTATATTAAACAAATGCTACAATATTA
TTAAATAAGACCCACACGCAATAATAATACAAATAAACATAATAAAACACAACTTTG
ACTAACCTCCGCGCTGTCCGAAC TCAATTACCCAAAATACAAGTACAGACACCAATCAA
ATTATAGTTGTGCGTCCGTATCGGGTCCACCTGTCCGATTATTAGTCGTCAGCATTGA
GCTCCTCTCTAAACAAAAACATTGTGAAACACTATTTGCCACTCAAAAAAGGATTTCTTA
TTATTTTTTTAAGCGAGTGCGCCAATAAGATCGGTCCGTGGGTGTCACACTGGACAAAGCA
GGCGGGCCACACTAATATAGGCAAATGCCACGCAGACGGCGGGCACACTAAGCCCAGCC
AGGAGCTATATAAGGCGGGCGTCTGTAAGGTATTAGTCACTCCGTATCCGACACCGATC
GGAGGAAGTATGCCTGCCGGCCGGGAAGGCGGTACGGTAGTGGTAGCTTCGGGGAAGG
CGCATTACGCACTTCCTGGTCAGTTGACGGTGTCCAGGACGACATCGTCGACTTCCATG
ACATCTATTAAAAAATTCGTGAAATCATGTCTCTGGCCAGATACAGCAGCAACAAC TA
GTCTGGTATGTGCAAGTCAATTTTCTTGTTCGAAACACGAAGTTAACGATCACAAGG
CATCACGCGCCACGATCTTACTGTCTTTTAATTTAAGGATGCGAGTTTGTCCGGCTTC

CCAACTATCAGCTTTTTTTTATTTTCACGAACTGGAACGGGATTAAAAATCCCCAAATAC
 ATTACCATTTTCAAAGCGGATTTTATACAACCGTATTTGGTATCCCAACGTTTGAAACC
 AAATACTGATAATGCTATAATACATAATGGTAATACAACCTACAGAGCAAACAGAAGTCC
 TGCCGTAGCTCTCGTATAAGTAGAACACTTTTTGTTTTTTGGTTCGACTATATATATAT
 CCGCACTATTTATTACAAAATAACTATCTTATATACCGAATATAATAAAAAAATTGGG
 ATGGCCAACCCCTTATTTTTTAATAACTTTTTTGGTGTGCGCAAGGTCCAAAGTGAGTAGAA
 AAGAAATGGCACTCTATTATTATTATTATTATCTATTATTATTTTCAAATATTCAATAC
 TTCTCGTCCGCTTATTTTGGTA AAAAGTGTGATATATAAAATTTCTAACGATTTGTCAA
 AAAACTTTTTTGGCACGCCCATTTTAACGCTCTAAAGTTCGGCAAACCTAATCACACCCACG
 CTTTGGACAAATTTTTAAATTTTTTCCATGTTATTTCTTAATATCTATCGATATCAGAG
 AAAAGTAATAAAATTTTCGCGTTCGTTTTTACACTAGCTGAGAAAACCTCGAACAAACAAA
 CAACAAATTATAAGGATGATGATACCGATTAAAGGCAATCTATTCGTGCTCCTCCTCCT
 CCAACAGATCGCTTGGCAATTCAGTGTGTTTGATCCAGATTCTTATGGCATTCTTTTTCG
 ATCTGACTGAATCGCTTTAAGTCTTCAGCATTCCTCCACGCTTTTGCCGCTTTTGATATC
 AACTCTTGTGGGTTTACGTCTCAATGCTTTTTTCCGGTTTTTTTTTACAGTGATGGAGTAT
 GTTTTTCCGCATGTGAATGTTGGCACCAGAATACGCAGGGGAAGAGAAGGCTTCTACAA
 GGATAAAAAAGTTAAGCCCATATTAGGTATAATTCAAATTTATCAAGTTTCTCGAAAT
 CCGAGTTAATATGGCACCAAGCTGATACCTTTGGGCTTGACAAAACATTAAATTTAAAAATAC
 TATATAAAATTTGAAATCTTCATCACCAGGAGTTCCTCTTTAGAATAGCTCGATATTTCA
 TTCGGCCAAACCGCGTGCATGATTTTATTTAAGTGCCCGCGTTTGCTGCCAGTGTTTC
 TTGCAGCATGTGAATGTGCTCATTACAAGGTGTGGGGTGGGCAAGGTGTTTGCTGTCA
 ATGGACTGCCCCCGGCATTAGTTTAAAGCCAGATAACGTTAGTTTGTAAAAATCGATCG
 TAATCCGTTTTTATTTCACTTACCCTCTGAATTAAGTTTTTTATTAAACTGAACTGAAGA
 AGAATTTTTTTTGAATATTTTTGTGCGTAGCTTAAAAATATAATATTATGGATATCGTTC
 CCTAATTATAATGTATAAAAAATGTACCGTGGAAAGATGCGCATCCTGGGAAACATCATT
 TTGTTAGGTGGCGCTGTAAATTTCTACATAGGTGGCAGAAATTATATGAACTATTTTTTA
 TTCCCATATTGCTCGTGGCCAAACTCTTTTTCGGCTAACAAAACCTCGGTAGAGCATAGATA
 TACTTTAAATGGTCGGAACCTACCTGTTACATACCTTTTCAAAGAGTGAGTATATACTT
 TCACTCTACGAGTAACGTGTATAACTACGTGTTTTAACCTTCTAACTTGCGTCGCCAGA
 CATTAGCTATACAAACGATTCTATTATTTTTATTACTATTATAGTTAAAAATATTTATAA
 TTGACAACACACTTTGACAAGATCGCTAAATAATTTATATTAAACAAATGCTACAATAT
 TATTAAATAAGACCCACACGCAAATAATAACAAATAAACATAATAAAACACAAACTT
 TGACTAACCTCCGCGCTGTCCGAACCTCATTTTACCCAATTACAACATGAAGAATTCAGA
 GTACTCTGAATTACCATTTGAAAATTCATTGCTTCATTCTTTGAACACATTCAGT
 TAGTTTATTTTCGATCGGACGCTGTAAAGTCCGAATACATACGAAGTGAACGCAAAGAGA
 ACGCGACCAACTGAATGGCATGTATTCTGAATACGCAGTAAACGAAATCGATCTGGAG
 AGTGGTGTTTTGATCCCTCTAGTACGTCTGGCCGTCGTTTTTCAATTTGTTGCTGCCG
 GGAAGGCACATTAGGCGCTACCAGCTCCTGCACACCGTTATCCAGGGGGAATTCCCTC
 CCCTCTGTGACAAAATTACAGCAACGTGTCATCAGGCCAAGGTACTGGAGCAGATCTTC
 GTCCGTTATTACATAAACAGTTCTTTATTAAGCAAAACGAAGTTCAATATCCCAAATT
 CGGCGTGAACCACGTTCTTTATGTGCTTCAAAGTAACGGGGCGGGGTTACCCCGGCTCC
 CCTACTATTATCAGTTTTTCTTGAATCACCGAATTGAAATTGGATTACGGTTCTTGACGA
 AAATATTATTTTGACTGACTTATTTATTTATTTCTTAAAGCACGTTTTTAATTTACCAA
 ACTTTAAAAGAAAATATCGAAAAAATACGTGTAAGTGTAACGTTAATATTGGTGAG
 GATTTTCAGAGAAAACCAAGGAGAAGTAATTAGGGCATTCAAAGTGATTAAAAGTTCTT
 TTCATAATTTCTTCCATCGGTGTCTACACCCAGTTTCACAAATAGGACAACTCGGCA
 TTAATTCGAGATTTTCCTTCTACGGTAACGGCGTAACGTGCTGTTGCACATTTACGG
 GTATCAATGG

***diver2* (delecionado)**

CGTGGGAATCAATGGCTGCATTCTGGAGCAGCGATATAGGACTTTTGAGGCCGTGGAT
 TGCGCCATGGCAACCTATGCGCCAGGCGTTTCAGGTGGGCAGAAATCGTTTCGACGCTAGT
 TGCTACCACCCAAAATTGTCTTGGTTGCATGCTTTGCCATAGTGCAGAGCATACCATAT
 ATTATTGCCCCCAATTTAGAGACTTATCGCCAGAAGATCGTCTGCGCGAGGCAAAGAGA
 CTAGCACTTTGCCTAAATTGCCTAAAGGCAGGTCATCAGCTACGGCAATGCAGCTCGAG
 CCGCTGCCGCACCTGTGGAATCAGGCATCATACGCTGCTCCATCTAGATGGTCCGCTT

CCTCGCAGCCACATGTTCCGGTGTCTTCAAGCTCCCTCACTGCCCCGCTTTTGAATTCT
TCTACTCTAATTGCCAGGATCTCGGTAGTGACCTTGTGCTGCTAGCCACTGCAACCGT
TCTAGTGCAAGATCGGTCAGGACTGTTTCGTTCCCTGCAGGGCCTTGTAGATTCTGGCT
CTCAACTGCACTGGGTACCTCTCGGCTTGCAAATCAACTGCAACTTAAGAGGTGGAGG
TCGTCCGGCTCCGTCACGTGTAATCGGGGATACCAATTTTCGCGACTGATGGATTCTCGGT
AGGAATTGCGATTTCGGTCTGTCACTTCGGATTTCTCAACGGCGCATTTTCTCAATGGTG
GTCCGCCTTCGTGTTTGACGAGCCAAATATAACGGGCCAACTATAATCGGCTTGAC
TCTTGGAACGCATCTCCTTTCTTCAGCAAATATTTTGGTCGCGATGGAAGGAAGAGTA
CTTGACGTTGCTCCAGCAGCGCTCCAAGTGGCGCACCCCAAAGCCTGGCGTAGCCGTGG
ACAACGTCGTTTCATGTTAAGGACGAGAATCTACCCCCAATGAGATGGCCTTTGGCGAGA
GTCATGCAAGTTGATTCTTGGCAGAGACGGCGTCGCTCGAGTTGCAGAATTGAGGACCGC
GTCTGGAGTAATAAGGCGGGCAGTGAACAAGCTGTGTCTGCTTCCCATTTGAGGACTCTG
TTGGAATCCAAGCTTCCAACGGGGGGAGGATGTTGGGTCATGCAGCCGCCAAAACGTG
CAGTAGACTAATCTCATACTTTTAAATATTGTAAGCCGAAGTAGTCAGTGGTAGCAGTT
GCGATGCCACAGGGCAAACCGCTGTTCTCGCACGCATTTATCTGTACGGCGCTCATTC
CTTGTCTTTTTGTTATCTACCTACGTTAAGCTTGGGCGCTGCATTTGGCTGTCTGTC
CCGCCAATTGCCACTTCTTCGTCTGTTCGGAAGAGACTAACTTGTG

-CG12717, CG12717, -2L35, -2L38, 2L38, CG12717, CG12717

TATAAGGACAACGCATAAGGACAACAAGGATAATATGCCCGCTCTACTCTATGAGGTGC
CGCAGCAGCAGAATCTCACAGATTGCGGCCTCTATCTGCTGCAGTATGTGGAGCAATTC
TTTACAAAACCCCTTCATCGACTATACACTGCCATTGGGGAGCTGAGCAATTGGTTTGA
CCTGCTCACAGTATTTTTATTGTTGGGGTAAGCGCGTTCCAAGTGCTGATCCAAAAT
TCATCTGTGCGAGTTGCTATCGCCATCGGAGTTCTGCTGCTCATATCGATACATATATG
TTTCCAATTGATATGGATTTTGTGGGTGTTTTTAAAAATATGTATCGATAGTCCGCTT
TGAAATTTGAATGTTGGGAATATTGAAAATAAATAATATTTGAACTGACCATTGTATCA
TTTACAAAATAGTGTTCTTACATGTGGTATTTACGAAGTTAAAGTTTCAGCCGTTGGT
TGCAAGTTCCTTTGTATGGGAATTTCAAAAACAAAATGTATTCTCGTATTTAATTTCAAC
AGACATTAATCAAATTATAGTTGTGCGTCCGTTATCGGGTCCACTTTTCCGATTTTTAG
TCGTCAGCGTTGTATTGTATTTGAACACATTTGTTCAATTCCTTGCTAATTTCCCGGAA
ATGTGCAAACGACAGTTACCGCCAGTTACAGTAGAAAGAAAATCTCGAATTAATTGTCC
TATTTGTGAAACTGGGGTGTAGACACCGATGGAAGAAATTATGAAAATAACTTTTAATC
ACTTTTGAATGCCCTAATTACTTCTCCTTGGTTTTCTCTGAAATCCTCACCAAATATTA
ACGTTTACACTTACAAGTATTTTTTTTCGATATTTTCTTTTAAAGTTTGGTAAATTAATA
ACGTGCTTTAAGAAATAAATAAATAAGTCAGTCAAATAATATTTTCGTACGGAACCGT
AATCCAATTTCAATTTCGGTGATTCAAGAAAACGTGATAATAGTAGGGGAGCCGGGTGAAC
CCCGCCCGTTACTTTGAACGACATAAAGAACGTGGTTACGCGCGAATTGTTGGATATTG
AACTTCGTTTTGCTTTAAGAAAGAACTGGTTTATGTAATACCGGACGAAGATGCTGTCCA
GTACCTTGGCCTGATGACACGTTGCTGTAATTTTGTACAGAGGGGATGGAATTTCCCC
TGGATAACGGTGTGCAGGAGCTGGTAGCGCCTAATGTGCCTTCCCCGGCAGCAACAAAT
TGAAAAACGACGGCCACACGTACTAGAGGGATGCATAACACCATTCTCCAGTATCGATT
CGTTTTACTGCGTATTGAGAATACATGCCATTGAGTTGGTCGCGTTCTCTTTGCGTTCA
CTTCGTATGTATTCGGACTTACAGGCGTCCGATCGAAATAAACTAACTGAATGTGTTCA
AAGAATGAATGAAGCGAATGAATTTTCAATGGTAATTCAGAGTACTCTGAATTCCTTCAT
GTTGTAATTTGGGTAAAATGAGTTCCGACAGCGCGGAGTTAGTCAAAGTTTGAGTTTTTA
TTATGTTTATTTGTATTATTTAGCTGTGGGTCTTATTTAATAATATTGTAGCATTT
GTTAATATAAAATATTTAGCGATCTTGTCAAATTTGTGTTGTCAATTTAATAATTTTA
ACTATAATAGTAATAAAAATAATAGAATCGTTTGTATAGCTTCTGGCGACGCAAGTTAG
AAGGTTAAAACACGTAGTTATACACGATACTCGTAGAGTGAAAGTATATACTACACTCT
TTGAAAAGTATGTAACAGGTAGTTTCCGACCATTAAAGTATATCTATGCTCTACCGAG
TTTGTTAGCCGAAAAGAGTTTGGCACGAGCAATAGTGGGAATAAAATAGTTTCATATAA
TTTATGCCACCTATGTAGAATTTACGACGCAACCTAACAAAATGATGTTTCCAGGATG
CGCATCTCTCCACGGTACATTTTTTATACTAATATTATAATATTATCTTTAAGCTACGC
ACAAAAATATTACAAATAAAATTTCTTCTCAGTTCAGATTTTATAAAAACTAAATTCAG
AGGGTAAGTGAAGAGTGCAATTTGTGTGGACCAAAATTTCAATTATTTAGACCAACCGC
ATAATTATTTTCATTGGTGTGCCTTGTCTGCCAATCAGTTTAACTTTTCCATCACC

154

TTTTCGCCAACCGGTACCGGTGGCCTGGCATATGCTCTTCTTTAAAAATTCTCCTCCTCG
CTATCCTCATCCATCACTAGTTGGCCATTTAATGAGGTAACTTAATGACCGGCAAAAT
TTTGGCGTGCTGATTGCTTTTCATTTCATTAGCTTCTGGGTGAGATTAGCGATATCCTTAC
GGTCTTAGTTACTGTGAGCATGTCAAACCAATTGCTCAGCTCCCCAATGGGCAGTGTA
TAGTTGATGAAGGGTTTTGTAAAGAATTGCTCCACATACTGCAGCAGATAGAGGCCGCA
ATCTGTGAGATTCTGCTGTTGCGGCACCTCATAGAGTAGAGAGGGCATATTATCCTTGT
TGAAGATATGCGCCAGCGCATTTCGCATACTTTGCCTTGTATCACAGGTGAGATAATCCC
GCAATATGGCAATCGCTCGATGTCTGTCAGGTGACCGCCAACGAATCGAAAATAAGAATA
AATGGCTG

***diver2* (delecionado)**

TGTTTAGCTATTGAGTTTTTTAAAAGCTTTTGTGAATAATAAATAGCCGAGGAAAGTTT
GTTTATTTTACTTTACTGTTTATTTTCGCAATTGTTTACAGCAGCTAGGCCCGACAAAG
AGCTATATCGAATGCACAAGTTTAGTCTCTTCCGACAGACGAAGAAGTGCCAATTGGCG
GGACAGACAGCCGAAATGCAGCGCCCAAGCTTAACGTAGGTAGATAACAAAAAGAACAA
GGAATGAGCGCCGTACAGATAAATGCGTGCGAGAACAGCGGTTTGCCCTTTGGGCATCG
CAACTGCTACCACTGACTACTTCGGCTTACAATATTAAGTATGAGATTAGTCTACT
GCACAGTTTTGGCGGCTGCATGACCCAACATCCTCCCCCGTTGGAAGATTGGATTCCA
ACAGAGTCTCAATGGGAAGCAGACACAGCTTGTTCACTGCCCGCCTTATTACTCCAGA
CGCGGTCTCAATTCTGCAACTCGAGCGACGCCGTCTCTGCCAGGAATCTACTGCATGT
CCTCGCCAAAGGCCATCTCATTGGGGGTAGATTCTCGTCTTAAACAAGAACGACGTTGT
CCACGGCTACGCCAGGCTTTGGGGTGCGCCACTTGGAGCGCTGCTGGAGCAACGTCAAG
TACTCTTCCTTCCATCGCGACCAAAATATTTGCTGAAGAAAGGAGATGCGTTGCCAAGA
GTCAAGCCGATTATAGTTAAGGCCCGTTATATCTGGCTCGTCAAACGACGAAGGCGGAC
CACCATTGAGAAAAGGCGCCGTTGAGAAATCCGAAGTGACAGACCGAATCGCAATTCTT
ACAGAGAATCCATCAGTCGCGAAATTGGTATCCCCGATTACAGTGACGGAGCCGGACGA
CCTCCACCTCTTAAGTTGCAGTTGATTGCAAGCCGAGAGGTGACCAAGTGCAAGTTGAG
AGCCAGAATCTAACAAGCCCTGCAGGGAACGAACAGTCTTGACCGATTCTGACACTAG
AACGGTTGCAGTGGCTA

CG12717

TTCAAAAATCATCTGTGCGAGTTGCTATCGCCATCGGAGTTCTGCTGCTCATATCGATA
CATATATGTTTCCAATTGATATGGATTTTGTGGGTGTTTTTAAAAATATGTATCGATA
GTCCGCTTTGAAATTTGAATGTTGGGAATATTGAAAATAAATAATATTTGAACTGACCA
TTTGATCATTTACAAAATTAGTATTCTTACAATATGGTATTTACGAAGTTAAAAAGTTCA
GCCGTTTGTGCAAGTTCTATGTCTGGATTTTTCAAAAACAAAAATGTATTCTCGTATT
AATTTACAGCATTTTGCAAATACAAGTACAGACATTAATCAAATTATAGTTGTGCGTCCG
TTATCGGGTCCACTTGTCCGATTTTTTAGTCGTGAGCGTTGTATTGTATTGAAACACATT
TGTTCAATTCCTTTGCTAATTTCCCGTAAATGTGCAAACGACAGTTACCGCCAGTTACCG
TATAAAAAAACTCGAATTAATACCGAGGTTGTCTATTGTGAAACTGGGGTGAGAC
ACTCACTTTTAATCACTTTAATCACTAATCACTTTTTAATCAACTTTAATCACTTTTG
AATGCCCTAATTACTTCTCTTGGTTTTCTCTGAAATCCTCACCATAATTAATCTTTGG
TTTAAGAAATAAATCAGTCAAAATAATATTTTCGTGAGGAACCGTAATACAATTTCAAT
TCGGTGATTCAAGAAAACGATAATAGTAGGGGAGCCGGGTGAACCCCGCCCCGTTACT
TTGAACGACATAAAGAACGCAGCGTACTAGAGGGATGCATAACACTATTCTCCAGTATC
GATTCGTTTTACTGCGTATTGAGAATACATGCCATTCAAGTTGGTCGCGTTCTCTTTGCG
TTCATTTCGTATGTATTCGGACTAACAGGCGTCCGATCGAAATAAACTGAATGTG
TTCTAAGAATGAATGAAGCGAATGAATTTCAATGGTAATTCAGAGTACTCTGAATTTCT
TCATGTTGTAATTGGGTAAAATGAGTTCCGACAGCGCGGAGGTTAGTCAAAGTTTGTGT
TTTATTATGTTTATTTGTATTATTATTTGCGTGTATGTCTTATTTAATAATATTGTAGC
ATTTGTTTAAATAAAATTTAGCGATCTTGTCAAATTGTGTGTCAATTATAAATAT
TTTAAATATAATAGTAATAAAAAATAAGAAATCGTTTGTATAGCTAATGTCTGGCGACG
CAAGTTAGAAGGTTAAAACACGTAGTTATACAAGTTAACACGGGCCAACGAGCAAGACC
AGTTGGCTCAGTGACGAAATTTAATACCATTACGTTAAAGTGCTGACGCCAGCATCGG
GCCGGAAACCCGTGCGCAGCCGGCAAACACAGCATATTTGCGTGGCCGCTGCGAGTCTT

TTGTTAGCTGTAAGATTTCATTTTTTATATTTCAGTTTATGTTTTACTTTAATAAAGAGCG
 GTTGCTCGTCTACATTATCACTCCGACTGTTTCGTCGCAAAACATACTTGTATTAATTAC
 TCTACCAGACCACAAGGCTGGTAGTTAAAAGAGGAAAGTTCCTCTCAACCTAATATATT
 GGACAGCAGCGAGGGATCCCCCTGCCACCCCAATATGAATATCTGGCGCCCAACCAGT
 GGTAAGAACTACTACCACCACCCACACAGCTACTTAATTTGCGGCAAAAGCAAAAGACAG
 CGGCGCGGGAAGTGAAGGGAATTAGTGACAAAATAACACACACACCAAAAAACAAAA
 ATTATG

gypsy6

TGAGTGGGAACCGATGCTAAACGCATTTTAAAGTCTGTGAATATTTAAAAAATATATA
 ATAATAAAGAATTATTAACAAAAACAAAAATCGTATAATTGTCTTTTATAACGGCGCA
 TACGTTTCGCTTTTCATTTATTGTTGTTGTTTGTGTTTCCATTTCAGTTCAAGCGCGACA
 TTCGACGCTGCAGCCTTTGTCTTTAGTTGTTGTTTGTGTTTGGCTTTGGTCAACTGCTGCC
 CACACAGGCAGCTTTTGCATACCTAAGCGAATTTCTCAGGAATAGGAATAGTACACCAT
 ACTCCACGGACTCAGAGTCCGATTGTGAAAATCAGCCTTCCCCCTAAGACGTGCAACC
 GCGAATTTCCCTACAATGGCGATAGACCCAGAACAACCTCAAAGCTGTAATACAGGCAGT
 GGTAGCCAATACTTTAGCGGACGAGGCTGCCAGAAACGAGCTCGCACAAAATGAAATGC
 GTCAAAAAATTGAGGATCTGGCCAGCCAGTTGGCAGCAACACGGGTTCGCGCCACCGCA
 GCAGTGGCCCCCGCAATCAAGGCTTACCTACCCGTAGACATAACTAGTAGCGTACCATG
 TGACAGGACTTTAGACGCCGTTAAATATTTGCCGGAATTCTCGGGGTTCACAGGAATCAT
 ATGTTTCGTGGCGACAAGCGCGGTTGCTGCGTATCGCATATTTAGAGATTACGACGGC
 ACCTCTCGCCATTACGAGGCAGTGACAATTATTAGGAACAAAATTAGGGGCGCGGCCAA
 TAGCGTTTATCCTCGTTTGGCACCGTACTGAATTTTCGATGCCATTATAAATCGGCTCG
 ATTTACGTTACAGTGACAAACGCCCAATGCATGTCATTGAGCAGGACATGAGCACTTTA
 AGACAGGCAAAACAACATGACCCTGTTGGAGTATTATGATGAGGTTGAGAAAAAATCAC
 CTTACTCACAAATAAAGGCCATATGTCTCACGAGCCATCGGCGGCTAAGATATTGTGCG
 AAAAAATCCGAGAGCAGCGCTTGCCTGTTTTTCATCTCGGGACTTAAACGCAACCTCACT
 GACGTGCTTTTACGCGGCAAGCCGAAAGACATGCCGTGAGCCTTGGCCTTAGCGCAAGA
 AGTGGAATCTTAACCACGAGAGGTACGTTTTTCGCGAGCTAGTTTCGCAAGAAGTCAAGAAG
 ACAAAGATCGCAAACTGCTGCAAAAGCGCAGGGTTCGTCAACTCGACAAGTACGACCGC
 CATGACCCACAACCTAGTAATACCAAAAACCCGTATTTTAATAGGCAGCATAGGGCACA
 GGTCCACGCGCCGACCCGCAAGGTACCAAAATTACAAAGACGACGGCCACCGGAGCCTA
 TGGAGGTGATCCCTCACTTTCTAAGTTGATGCAGCCGACCCAGGCAAAATACCTACCGG
 AACAGACGACCAGCCACCTCCGAGCGCACAGGCGCCGTGAGGAAACAACAAAAAGTTAA
 CTTTATTACGAGAATGCGGAAGAAAAGACAGGAGCATATGCCGCCGCCGCCGCAAAAG
 CGGAATCGAAAAATAGACGACGATGCCATTACCGAGTATGACTCTGACTACCTCAATTTT
 TTAGGGGTAAATCCCTGTGTACCCGTATCAGACGAAGAGTAGCAGGGGTACCAATGAAA
 TTCTCTTATAGACACCGGCGCATCAAAAAATTTTCATCCGGCCTCGCAAGAGCTAAAGGG
 TGCTGCCCCGGTGGACTCCCCTTTGAAATTCATTCAATTCACGGCACCACCATCGTTA
 CAAAAAATGCTTCGTCTCGATTTTTTAATTTGAAAGCGACTTTTTTTCATTCTACCAGAT
 CTTACAACATTTCGACGGTATAATCGGGGCCGACCTGTTAACACAGGCCGGTGCATCACT
 TTGCCCTCGCCTCCGCCCAACTCAAATGGGGTCAGGAAGTTGAGAAGATTTTCATTCCATA
 AATGCACTGACGTCAACTTCACTAACGTGGAGTGCGCAGATGCACCACCTTTGGTGAAG
 AAGGCATTTCTAGGGATGATAAGGAGCCGAAAAAACGCCCTTCGCAGATCCCAACGAGGC
 CCTGCCATATAACACATCCGTTAGTGGCCACGATTAGGACTACTAGCGAAGAACCCATTT
 ACGCTAAATTTATATCCATACCCCATGGGGGCAGCAGATTTTCGTCAACAACGAAATCCAA
 AACCTGCTTAAAAATGGTATAAATCAAAAGTCGGTGTCTCCGTATAACAACCCAATATG
 GGTGGTTGATAAAAAAGGAACCGACGAGGCCGCAATCAAAATAGACGATTAGTCATAG
 ACTTTCGTAAGCTAAACGAAAGAACTATCCAGACGAATACCCATATGCCAGATATCTCC
 ATGATACTGAGCAACTTGGGCAAGGCCAAATTCCTTACGACATTTGGACCTCAAGTCAGG
 GTATCACCAGATCATCTTAGCGGAAAATGACCGCGAAAAAATTCCTTCTCCGTAAGCG
 GGGGGAAGTACGAATTCAAGAGGCTTCCCTTTGGCTTGAGAAATGCTGCCAGCATCTTC
 CAGAGAGCCATTGATGACATTCTCAGAGAGCAATCGGCAAGACTTGCTATGTCTATGT
 CGACGATGTAATAATCTTCTCAGAAAACGAGAATGCTCATGTCAAGCACGTGGATTGGG
 TTTTAAAAACCATAAGCGATGCAAAATATGAGAGTCTCAGTGAAAAAGTCAAGTTTTTTT
 AAAAAAGCGTAAACTTTCTTGCTTTTATAGTCACCAGTGATGGCACTACCCTGACCC

AGAAAAAGGTCAGGGCCATTAAAGAGTTCCCTGAACCAAAAACGGTGTTTGAAGGTTAGAT
 CATTTCAGGTCTCGCGAGCTACTACAGATGCTTCATTAAGGACTTCGCAGCCATAGCA
 AGGCCTATTTTCGGACATCTTAAAGGGCGAAAACGGAAGTGTAAAGCAGACACAGATCGCG
 ACACGTCCAAGTACAATTCAATGAGGCGCAAAAAAATTC'TTTTGAAAACTGCGCAACA
 TTTTAGCATCCGAAGATGTTATGCTCCGATACCCGGATTACAAGAAGCCATTCCCATCTA
 ACGACGGATGCTTCAGCCTACGGTATTGGAGCAGTGCTTTCACAGGAAAAACCGTCCTAT
 TACAATGATCTCGAGGACATTAAAGGACAGGGAATGAACTACGCCACAAACGAAAGGG
 AATTATTTGGCCATCGTTTGGGCTTTGGCCAAACTGAGGCACTATTTATATGCGGTGAAA
 GATATAACTATCTTTCACCGACCATCAACCATTGTCATTC'TCGGTATCAGACTCTAACCC
 TAACGCAAAAAATTAAAAAGATGGAAGGGTCGCATCGAGGAAACGGGTGCGAAGGTGGTCT
 ATAAACCGGGAAAAAGAAAATTTGGTTGCTGATGCCCTGTCTAGGCAGCAAAATCAATGCC
 ATAGAAGAACAGGACGCAGAATCATGTGGTGCGACCATTACAGTGAGATTTCCCTCAC
 TCACACCATAGAAACTACGGATAAGCCCCCTAAATTGCTTCCAGAACCACTAATTCTGG
 AAGAGGCCCGCTTTCCGCTAAAACGCTCTTTTCGTCTCTTTTCGAAATAAGAAACGACAT
 ACAATCAACTTCACTGACAAGGAATCATTACTCAATGACCTTGCGGACGCAATAGTCCC
 CAAGGGCGTAAACGCCCTCCATTGCGATTTGCACACGCTAGCAACGGTGCAGGACGACT
 TAGTCCGGAGATTTCCGACTACAAAATTTCTGGCACTGCAAAAACCGTGTTACAGATATT
 TTTGGGGTCGAGGAGAAAAAGGGAATCATACTGGCAGAACACAATAGGGCCCCACCGATC
 GGCCCAAGAAAAATGTGAAGCAGGTTCTCACAGAATACTACTTCCCAAAAAATGGCCAAAC
 TGCCCAATGAAATTGTGCAAAACTGCAAAACATGCGCTAAAGCAAAGTATGATAGGCAT
 CCTAAAAAGCAAGAGATTGGCGAGTCCCCGATTCCCTCTCATGTAGGAGAAATGCTACA
 CATAGACATTTTCTCAACAGACAAAAAATATTTCCCTCACTTGCATCGACAAGTTTTCGA
 AATTTCGCCGTCGTACAGCATGTACCGTCAAGAACAATTGAGGACTTGAAACCGGCCTTG
 TTACAGGTCATGAATTTTTTCCCAAAGGCCAAAGTGATTTACTGCGATAACGAGCCATC
 GTTAAAAATCGCACACGATCACGGCCATGCTTGACAACCATTTTGGCGTCAGCA

F

TCAGCGAAAGCAAAACCGCTAGCAATAAACGGTACTGCTGCACTGCCAGCAAAAACAAAA
 CGAAAACGTAAACAAAAAAGCTGGGTGACCTGGCAGACTGGAATGGACCGCTACATTA
 CAATAAAGCGAAAGCTCAGCCCGGAAAAATTCAGATTTGGGAAACAAGCCGAAAAATACA
 CGCGATAACTCTACCTTGATCAAAAATGTAGCCCCTGCAATACCAACAGATTTGCCTT
 GCTGGTAGATACCGCTGAGGACGTGCCGCTGGGATCCGTTGATATCGAACCGAAGAAAA
 CAAAGCCTCCGCCAATATACATCCGCGAGAAGAGCACAAAGCCGTCTTGTAATACTTTG
 ATTTGGCCTTATTGGGAAAGATAGCTTTTCATATAACTCCCCCTCGTAAGAGGTACTATCAA
 CGAAATCAAACCTCAGACGAAAACGGAGGACGACTACAGAAAAGTCACAACTATTTTA
 CCGCACAAAAAATAGGCTTCTACACCTACCAGCTTAAAAGCAGCAAGGGCCTGCAAGTA
 GTCTTGAAGGGCATTTGAGTCTGATGTTTACGCCCGAAGAGATAACTGAGGCGCTAAAGGA
 AAAGGGATTTTACGCCAAAAACGTGTTCAATATCAAAAACAGAAACAGGCAGCCCCAAC
 CACTCTTCAAGATTGAGCTTGAACCAGAAAACAAGCCTCCTAGAAAAAACGAGGTTTAC
 CCAATTTACAAACTCCAGCTCCTTTTGCACCGTAGGATCACGGTAGAAGAGCCGCACAA
 ACGCAACGCTCCTGTACAATGTACAAACTGCCAAGAGTATGGCCACACGAGGTCATATT
 GTACACTTCGCCCCGGTGTGCGTAGTCTGTGGAGATCTCCACGACTCCAAACAGTGTCAA
 ATTAACAAAAGAAAAATGCATGCGAGAAAAAATGTAATAACTGCGGGGGCAATCACACAGC
 AAATACAGAGGCTGTCCAATCTACAAAGAGCTGAAAATCCGTCTTCACAAAAGAATGA
 ACACGGCGCGGGCACACCAAGGATCAGCTACCCTGATACCATCAGAGACAAATCCTGAA
 GTAATTTTCTCGAAAGCAGCTAGTTTTCGCTCCCTGGCCTACATTCAACACTAACAAAGAC
 AACATTTTGTCTAACGTTTTTAAAAATCAGGTATGACGCCCTCCAACCCAAAACCTCCCGAATC
 CACATGAAGTGCACACAAAAATTAGACACACAACAAAACCTATCACCCAGCTGCGCAGCAG
 GAAACAAAAACTGAAGCTATGATGCAAGCCTTACAACAGAGCATGATGGAATTTATGAC
 ATTTATGAAGACCACCATTTCAAGACATGATGCGTAATCAAAACCTTTTGATACAAATGC
 TTGTAGCCCAACAATCAAATAAATAATGGCTACCTTACGCATAGCTACGTGGAACGCCA
 ATGGCGTCTCACAGCGCAAACTTGAGCTAGCTCAATTCCCTACATGAGAAGCATATCGAC
 GTAATGCTTCTTTTCGGAACTCATCTCACAAGCAAATACAAATTTTCAAAATAAGAGACTA
 CCATTTCTACGGTACAAATCATCCCGACGGAAGACACAGGTGGCACCGCCATACTCA
 TAAGGAACCGTATGAAGCACCACCTTTTACAAAGAATTTGCGGAAAAATCATCTTCAGGCC
 ACATCTATCAACATTCAGCTGGATGACAACACTCTCCTTACACTAGCGGCCGTATACTG

CCCCCCGTTTACAGTATTAGAAGCTCAATTCCCTGGATTCTTCCAAGCACTAGGGC
 CACACTTCATTGCAGCAGGCGACTACAACGCTAAACATACTCACTGGGGATCGCGACTT
 GTGAACCCAAAAGGAAAACAGCTTTATAAGACGATAATAAAAGCCACTAATAAACTTGA
 CCATGTTTCCCCCGGGAGTCCTACATACTGGCCATCAGACCTCAATAAGCTGCCAGACC
 TGATCGACTTCGCAGTTACGAAAAATATTTCCCGCAGTTTGGTTAAAAGCTGAATGTCTG
 CCGGATCTCTCATCTGATCACTCGCCTGTACTAATTCACCTCCGCCGATACGCAGAAAA
 CGTGAAACCACCAACCAGATTGACCTCTAGCAAAACAACTGGCTCAGGTATAAAAAAT
 ATATAAGTTCACATATTGAGCTAAGCCCCAAACTCAATACTGAATCTGATATAGAGAGC
 TGCACGTGTGCATTGCAATCATCCTTACTGCAGCAGCTCTTACTGCAACACCCAAAAAT
 AACAAATAATACAATTAATTCAAAAAAGACCAACGTACAAATCGAGCAACTCGTCCACG
 TAAAACGTCGCTTACGCAGAGAATGGCAATCTTCCAGATCCCCAACTGCAAAACAAAAG
 CTAAAAGTAGCCACACGGAACCTGGCCAACGCTCTGAAACAAGAAGAGGACGACGATCA
 GCGCGATACATAGAGCAACTCACACCAACAGGCACAAAACAAAAGTCACTGTGGCGAG
 CCCACTCAACTCTTCGCCACCGACTGAAACCGTTTTGCCGATAAGGAATTCATCAGGT
 GGCTGGGCCCCGTAGTGATGAAGACAGAGCCAACACATTTGCCGCTCACCTACAAAATGT
 GTTACGCCAAACAGGCTACTAGCACATTTCGCGCTACCGTCTATCCCGTAAACCGCC
 ATCAGCAACACACCCCAATTGTGTTTCGTCTTAAAGAAATACTAAAAATATTAAGAC
 AATCTCAGCCCCGAAAAAATCCCCGGCTACGACCTTATAACACCGGAAATGATCATCCA
 GCTGCCACATTCTGCAGTTTCGTACATAACCAAGCTCTTTAATGCCATCACCAAACTTG
 GTTACTTTCCACAACGATGGAAGATGATGAAGATCATAATGATTCCAAAAGCCTGGTAAG
 AACCACACAGTCGCTTCATCTTACAGACCAATAAGTCTACTCTCATGCATTTCGAACT
 ATTCGAAAAATGCCTGCTGATCCGACTTAATCAACATCTGATATACCACAATATAATCC
 CAGCCCACCAATTTGGATTTTCGCGAAAGCCACGGAACCATTTGAACAGGTGAATCGTATT
 ACAACGGAAATAAGAACTGCATTTGAATATCGCGAATACTGTACAGCAGTATTTTTAGA
 CGTATCCCAAGCATTTCGACAAAGTCTGGCTCGACGGCCTAATGTTTAAAATTAATAAT
 CCCTACCCGAAAGCACACACAACTTCTAAAGTCTTACCTCTATGACAGAAAGTTTGCA
 GTGCGGTGCAACACTGCCACTTCCACTGTTTCATACAATTGAGGCTGGAGTCCCCCAAGG
 CAGCGTTCTTGGGCCAACGTTATACCTCATCTATACAGCCGACATCCCTACAAATAGTC
 GCTTAACGGTATCCACATTTGCCGACGATACAGCTATCCTTAGCCGTTCAAGGTCCCCT
 ATCCAAGCTACAGCACAGTTGGCACTGTACCTCATCGACATTGAGAAGTGGCTCTCTGA
 CTGGCGAATAAAAGTAAACGAGCAAAAATGCAAGCACGTGACGTTTACGCTAAACAGAC
 AAGACTGTCTCCGCTCTTGTTGAACAGCATACCACTCCCGAAAGCAGACGAGGTAAAG
 TACCTAGGAGTACACCTAGACAGAAGACTCACATGGCGCAGGCACATTGAAGCCAAAAA
 AACCCAACCTAAACTCAAAGCCAACAACCTTACACTGGCTCATCAACTCTGGTCTCCGC
 TCAGCCTAGATCACAAGGTCTTGCTCTACAATTCTATATTGAAACCAATCTGGACCTAT
 GGCTCACAGTTATGGGGCAATGCCAGCAACAGCAATATTGACATCATTCAGCGAGCACA
 ATCAAAGATTCTGAGAACCATCACTGGGGCACCGTGGTACGTTTCGGAGTGAAAAACATCC
 AAAGAGACTTAAATATCCCATCAGTTACCAACGCAATCACGGAACCTAAGGAAAAATAC
 CATAGCAAGCTTCACACGCACCCCAACCACCTAGCGCGAGGTCTAATCCAGCTCAGCAG
 CCGTTCCCGTCTCCGGCGAAAGGACCTACCAACCCAGCGAATAAATTATTAGGGCCGTT
 TAAACATAGAACAGTTGGAAAAATAATACAACCTGTTCAAAAAATACTGTTATAGTTAA
 GATTTTTAAACTTATTGTTAGTTCTTATACAAGAAGATTCAATAAATAAAAGCAAAAGTA
 AAAAAAAAAAAAAAAAAAAAAA

gypsy6

GCATCACGAATGCGCCGCCCTTACAGTGTGTCAAACGGGCAAGTGGAACGTTTTTCAC
 AGCACCTTACTAGAGCTCGCCAGGTGCCTAAAAATTGACAAGGGCATGAGCGACACCGT
 GGAGATAATTCTGTTAGCCACAACCAATATAACAAATCAATCCATTTCGGTTCATCGATA
 AGAGGCCAGTCGACGCAGTACAAGAGTGCACGGACGACACAGAGAACGGATTGTTGAC
 AAAATTAAAGAGTGCCCAAGACGCGCTAAGGTCTAGAGAGAATGTTTCTCGGCAAAATAG
 AGTCTTCGAAGTGGTCAAAAAAGTTTTGGTCAAACTAACAGAAGACTTGGCAATAAAC
 TCACTCCCTTGTGTGAGGAAAAAGCCATACAAGCAGACCTGGGGACCACGGTCCCTCATT
 GAAGGGAGGGTGGTCCAAAAGGACAATTTAAAATGACTCTCTCCCTTAATTTTTAATTT
 TTTATTATCACTACCTATTATCAGCCGTTTGGCATACTTCGATTTAGTAGATATTAGGT
 CCGTTGCCAAAGGCGCGACGTTGTAGTTTTAATTGGTATTAAATTTAGTATTAAATGGTG
 ATGGGTTAAACCTCGGCGAACAAGAATAAGCGACTAAAAAATACATTTCCACAGGTGCA

GAATTTTCATCCTCCTCTCACTGGCATTGGTGTGACGCGCGCTCACTGACTATTTCGCAA
 ACAAGGTATATCCCCATCATAGATGGAGAAATCCTGGTATGGGAAGAACTCGCATATGT
 GACCCACTCAGCAAATCTCTCGGAATACATGCGCATAGTCGAAGAAAACAAGCAGCATGA
 ACGAGATGTTTCCGAGTCTCATATGAGGAAATTGCTAGACGTGGATACCTCGCACCTT
 CGGGATACGCTGGATTTCGCTAAAAGTCCACCACAGAATAGCAAGGAGTTTAGATTTTCT
 AGGCTCAATGCTAAAGGTAGTAGCGGGGACGTTTACAGAATGGCAGTTGACACAGTCAA
 ACAATAGGCAGATCAAAATTAACACTAGGGTACAAAATCGGATTAACCAATTAACAACA
 ACAGTAAATCAAATCCTGCAAAACACAAAAATATCCAAATTGACACCGCCATTTATA
 CGAGACACTGCTGGCTAGGAATAGAATGCTAATGATGGAGTTACAGAACTTAATGTTGG
 CTGTAACACTGGCAAAAAACAATATTGTGACGCCAAATATCCTAGATCATGCAGACTTA
 AATTGAGTTTGGCTGAAAGAACCCACCGATACCCCATAGGGGATCTTATGTCCGCATC
 GTCTGTAAAAATACTGCAATCCCTAACATGATACACTTTATTATCAAAATTTCCAGAA
 TAAAATTATCTTGCAAAAAGGTCACTATTTTCCCAGTCACTAGCAAAGGAGTTATGCTG
 CGGATAACCGACAACATGGTAGCAAAGTGTGGTGAAGTAGTCCACACAATCAAAAATTG
 CATCCCAACATCGGGGGCTACCTTTTGCCAACATCAACGGAAAGCTCATGCGCCAGGG
 AACTACATGCGGGCGTCTTAGCACATTGCGAGTCACAACCAAGTGACCTACACCCGCTC
 ACCCGCTGGACGAGGGTATCATCATCAATGACCGGCCAGCCAGGGTCACGGTAGA
 CAACGAAACGGAAGTCTACCTACGCGGCACACATCTCATCACCTTCAACGATCGCGTCG
 TGATAAATGACACCACCTTTTGTAAATCATGACAAGGCCAAACAAGAGCTCCAGGGGTA
 GCGAGTTCCCCATCATTGAATATCACCGCCAACAGAGATATTCTGAGCCTCCCTACCT
 TCACCAGCTAAGTGAACGTAACCTTGGAGTTCATCAAAGAGTTCCGGGAGAGAAATTGATA
 CTAATCATCATCATCGCATAATATTTATCGCAGCAGCGATTTGTTGTGCAATTGGTTTGC
 ATCGGCATCACCTGTCTACGGTACATCGGAGCGCGGAGATCTGCAATCCAGTTAAACGG
 GATGATCGCTGAATTAGGACCAACCGAGGACGGCCGCAATCTTGAAGGGGGGATAGTTA
 ACACGGGCCAACGAGCTAGACCAGTTGGCTCAGTGCACGAAATTTAATACCATTACGTT
 AAAGTGCTGACCCAGCATCGGGCCGGAACCCGTGCGCAGCCGGCAACACAGCATAT

-roo

TCAAAGTTGTTCTGGTTATCTAACAATTTCTTCATATTTTCTTCTAATTGCTCTCTATC
 ATCTTCATCCATTATACCAAATAAAATATGATACAAGGAACCCATAAATTCGAAAGGAG
 CACGCTTGCTCTAGATCTAGACTGCATCATAAACAATTTATTGTTTTCTTCAAGTTCC
 GATAACTGACTTTGCATATTATCTAAGACTAGACTACATTGCTCTTCAAAGCTATGAAG
 TCTTTCGCAAACTTTCTCATACTTTGTATAAGCGCATTACCCTTTGTTAAACATTTTAA
 AATATGGATCCATTTTATAATAGATAACCAAATTTCCAAGAAGTACTCACAACTCAACA
 TCTCCTAGCGGGTCTAGATATATTGCTGAGGTTTTATTTATTTTGTCTATAGAATATCT
 TGGTGCTATATCTTTAGGTAATGCGCTAGAACTTGACAACCTAACACAAACAACAACA
 TTGCCATGATGATTCGATCTTGGACATTTCCCGATGTGCTCTATTTTCGCTTTTTTGGC
 TCTTGGTCAGCCTCATTTTTTGTCAACAGACTTTATTCCTTCCAAGGGACAAATTTAGT
 AACGGGTCTAGTGATATATCCTTCCTGCATCTTTACTTTAGCCACCCGGACCTTATCAT
 CATTCCTTATGGACCTTTTCCACCTTTTCTAAAGGCCATCTTGCAAGGATGACAATTC
 TCATCCTTTAATAAAACTATTTGCCCTTCTTCTATATTAGGAATTTTCTTTTCCATTT
 ATTCCTTTGCTGAAGCGTATGCAAATATTCATTTTCCACTTAACCCAGAAATCTTTCC
 TCATTTTTTTTGATAAGTCTCCACCTATCCAAATTTCCGATTTTTTTCATCTTCCATTGGT
 TCGACTATTTCTAAAGGTGGTCTTCCAATTAAAAAATGACCTGGTGTTAAACCTCTTG
 TTGGTCCATCTCACTAATACTATAGTGATAATGGCCTTGAATTTAAGCATGCTTCTATTT
 GACATAAAAGAGTTGACATTTCTTCGTAAGTCAAAATAGTGTGCGCGATTATACGCTTT
 AAATGGTATTTTCATTGACTTAACCCAGCTTCCCAAATACCTCCGAAGTGAGGTCTGTC
 CGGGGGAATAAAATGCCAATCAATCCTGTCTTTTCAAGCTGCGCTGCAATCGTTATAT
 TTTCTTGTATTGCATTAAATAACTCTTGATCTAATTTTCTTGGCGCTCTTACAAAATTT
 GTTCCGTGTCTGAATAGATATTGGAACATTTTCCCGCTTAGCAATAAATCTTCTGAG
 TGTCTGTAATAATGCGTCAGAAGTTAGATCGCTTACCATTTCTAAGTGATGGCTTTGG
 TGGCCATGCAACGAATACGGCAACGTATCCTTTAAATGTTTTTGGCCACGATTTTTT
 GAACATTTAACAATAAAGGACCTGCGTAATCTATTCAGTATTAAGAAACGGGAATGT
 CATCGTCACTCTATATTTTGGCAAGTTACCCATTATTTGCTGAGCTGTATTTTGTATTAT
 ACGTTGCACACTTTACACATTCTCTTAAATACTTCTTCAACGAATTTTCAACCCGAAA
 ATCCAATACTTTCTTTGGATATAGTTTCGCATAAGGTTTATCCCTCCATGCAATGTTTC

CTTATGAGCATT TTTTATTAATAAGCTTGT TAGGTGGCATT TTTTCTAAAATGATTGGAT
 GTTTAACATTAAAT TCTGCATTGGAGTTT TGAATCTTCCTCCAACCTCTAGAACCCCA
 TCCTTGTCAAAAATGGATTCAATGACAATATTTTATTATTGTCTTGATTTCCTTTTT
 GATTTTAAGGCACCTTTATCTCTTGCTAAACTGGTATTCTTGTTGTTTCTTAATAACAA
 CTGTTTCCGCTATTCTTATCTCCTTTACTGAAATAATTGATGAATAGGCTTTATTTTTT
 GTTTTCATCTGCACGAATCTATTTATGTATGCTATTATACGTATAAGTTTTCTATACT
 GGAATACCTTTCTATTAATTCGTAAATAGGATCATCTATTTTGTGATGTAATACCGTAT
 TTATTAAGACAGGTTCTTCTACAGACTGCTGCCGAGGCCAAAGTCTTTTGGGTCGGCT
 AGCCATTTTCGGACCTTTCCACCAAAAATCACAGTTGATCAACTGGTTAGAATCCCACTCC
 CCTGGATGCTAAATCTGCTGGATTATCCTCTGACTTAACATGATTCCATTCACTATTTTT
 TTAATTTCCGAATGTCATCCGTTCTTCTTCTTATAAATTTGATCTTACTTTGACCACTG
 TTAATCCATGCTAAGGTAATCGTGGAATCACTGCAAGCATAGATCTCCATTATATTGTC
 AATTGATCCTTTTAGTCTTTGGATTAATTCAC TAAGCAGGTGAGCTGCAATTTGGGAAT
 TGGGAATTGTCTTTCTATTTTTATAGGGTTGACTCTACTTTTGTAGCTATTATATTAA
 CATGAGGTCTTACTTTAGCATAGACTACTGCAGCATATGCTTATTTCGGAGGCGTCCGCA
 AATCCGTGAATCTGAATGACTGAAGAACTGTTTGAATTAATCCACCTTGGGATTGCAAT
 ATTTTCTACAATAATAAATTTTCTTTTATATTATTTTTCACAATAAATTTTATCTTCTATGG
 ATAAATCCTGATCCCATCTTCACTTTTATTTATCCAAAGTTTTTGAATAAAAAGTTTTCTCT
 GAAACGGTGACTGGTGCCAACCATCCTAACGGATCAAATATTTTGTAGCGTCGATAA
 CACAACGCGCTTATTTATATTTTTTGATTCAATTACAATTTACGCTGAATTTAAATAAAT
 CATTTTGAGGTTCCCATTTTAGTCCTAAAGTTTTAACACATTCATTTTCGATAATATTG
 AGAACCTTATTGTCCCTGTGTCTCCACAGTGGTTAATATTTTGGAAATGTTGGAAAT
 CCATTTCTTAAAGTTGAATCCAACCTTTCTGCAATTCATGGGAAATTAATGTTATTAATT
 TATTAGCTTCTTCTACCGAATCAGCTCGAGTCATTAGGTCATCCATATAGAAATCATTC
 CTAATTATTGCACTAACTTGGATTTTACATTTATCTGCAATATCTACCAGAACCCT
 GGTAGCCAAATATGGTGACAGACGAGTTCGGTAAGTGACTGTGGTTAATTTATATGTTT
 TAATTTTTCTTTTGGAGAATTTCTCCATAAAAATATATTGATATTTTGTATCATTTATTA
 TCTATTTTAATTTGTGCGGTACATCTTTTCAATGTCTGCCGAAACCACAAATTTCCCATTT
 TCTCCATTTAATAAATGTCAAAAATATCTTTTTGAACTCGTGGCCCAACCCACATTA
 TGTCGTTCAAACCTTTTGTTATTCGTAGTTTTTGTGTAAGCATCAAAAAC TACTCTCAAT
 TTGGTCGTAAGGCTTGAATCTCTAATCACTGCCTGGTGCGGTAAAAAATATTTGCCTTC
 ATCACTCACTTCAATCATGTGTCTTAAATCCATGTATTCATTCATGAATTTAGTGAGT
 CAACCTTAAGTTTTTCATTTCTTTTGTAGTTTTTCTCCAGATTCATGTAACGAGCTATC
 GATTGTTTCTTTGAATCTCCTAAGGTGACATCCTCCTTGAATGGAATGACACAATGTA
 TCGCCCATCTGAATCTTTTTTGTCTTTTGTATAAATTTATTTTACAGATTTTCACT
 CGATATCATCTTTTTCTTCTTCTTCCACTTCCAGTATCTAACTCTTTTATTTCTATTG
 TTGTGGCTACAATGGTTTCTTTTCTTTTGGATTTTTTACATCCAGAACTATCCACCCG
 AAATCAGTTTTTTGCCAAGGAGACCATCTATTTTTATAACTCCATTTTGCAGAATGTG
 AGTATATACGCTCTGCTCCAATGATTAGATCAATACGACCCGGTTTATTTAAATCGGGT
 CGGCTAATTTAAAGTTCTTCCATTTTTTCTGATCAACATTAATCGTGTGACTGGAAGT
 GCCTTCATAAGCTTTGGGAGAATAAGTGCTTCAATTTCTAAATTTTTCGGAGAATTTCT
 TATCGAAATAACCGCTTTTTGCTTGGAGATGCACGTTCCGTGGAAGATACTCCACTTA
 TTTCACTATGAGACCGAAATTTTTTCAATTTTAGAATCTGTGCAGACTCTTCTGATATA
 ATTGTGCTTTGAGAGCCACTATCAATCAATGCTCTTAATTTGTTCAAAGCCTCCATACCT
 TGACTTTACTTGAATCAAGGCCGTGGCCAACAAGGCTTGACCTGTTGTTCTACATGTAT
 TCACTTTTTCTGGAATTATGACCTACAAAGTGAAGTAAAGTATGGTGAGGTTTACGACAA
 GTCGAACAAAGCTGCTCGCTTATACATTTTTTTTCCAAACGGATGCCTCAGACATCTTAG
 GCAAATCCCATTTTTTCTTACCCAGTCAGACCGTTCTGCTGGATTCAATTATTTAAATT
 TATGGCATTGAATTAATAATGCCCTGGTATTTTGAATATGCACAATTGTCACTATAA
 TTTTATTCTTGTTATTATTAATCATTTTCTTTACAGGTTTACTTCTGTGAAAATGA
 TGATATAGAATTTTGTCTTAAAGAATAGAATCTTTGTCTTAAAAAGTCCATGACATCAG
 AAAGTGCTGTATTTCTTTTGTCTTTTTAAATTGAGTGATTCTTTATGAAATTTCCGAA
 GAATTAAGTGAGCGAAAATTGCTATCCACATCTTCTGGTAATTGTGCTTTAATTTTATG
 ATATAAATTGACTCGTTAATCGTGTCAATAAATGTTTTTACTTGCTTATTGGATTCCAA
 ATTTAAATTTTGGCATATCCATAAGCCTATTCATATGATCTGAGAATATGTTTCTTTTAT
 TCTCATACCGCTTGGTCAAAAAC TCCAAGTGGCTTCATAATTTTCTCCAGAGCCGAGC
 AGTAAATGAGTAACCACATTTCTGGCTTCTCCTTTTAATGCTGACTTTAGATAATTAAA

-2L35, 2L38, -CG12717, CG12717

161

GCGGATATATATATAGTCGAACCATAAAACAAAAAGTGTTC TACTTATACGAGAGCTAG
 GGCAGGACTTCTGTTTGTCTGTAGTTGTATTAGCATTATGTATTATAGCATTATCAGT
 ATTTGGTTTCAAACGTTGGGATACCAAATACGGTACTAACTAATTTATTTATTTATTA
 TTTATACCAGGAAGTGCCTAATGCGCCTTCCCCGAAGCTACCACTACCAATCGGTGTCTG
 GATACGGAGTGACTAATAGCTGACAGTCGCCCCGCTTATATAGCTCCTGGCTGGGCTTA
 GTGTGCCCCGCTGTGCGTGGCATTGTCCTATATTAGTGTGGCCCCGCTGCTTTGTCCA
 GTGTGACCCACCGACCGATCTTATTGGCGCACTCGCTTAAAAAATAATAGGAAAGCCT
 TTTGTAATAGTTACCTAATTTGATAGTGAGAATTTACTGCGCGACCGAGTAGCAAAT
 AGGGCTTCACAATGTTTTTTGTTTAAAGAGGAGCTCAATGCTGACGACTAAAAATCGGAC
 AGGTGGACCCGATGACGGACGCACAACCTATAATTTGATTAATGTCTGTTGAAATTAATA
 CGAGAATACATTTTTGTTTTTTGAAATTTCCCATACAAAGAACTTTCAACAAACGGCTGA
 ACTTTTAACTTCGTAAATACCACATGTAAGAATACTAATTTTGTAATGAAAAGTAAAG
 AACTTAGCTTAGCTTAGACTCGCTTAAATGCCTATATGTGGCCAGTTCAAATATTATTT
 AATTTAATATACGCAACATGCAAATTTCAAAGCGGACTATCGATACCTATTTTTAAAAA
 CACCCACAAAAATCCATATCAATTGGAACATATATGTATCGATATGAGCAGCAGAACT
 CCGATGGCGATAGCAACTCGCACAGATGATTTTTGGATACAGCACTTGGAAACGCCGTTT
 ACCCAACAATAAAAAATAGAGAGATATTAATATCGTATAAATCACACAGATAGGAAAA
 TGAAAGGAATGTAGTGATCAAAAGTCACCGTGTATGGTGAAGTGTCAACCATCCAGTTAA
 CTTTGCACATCAAGCAGACCGAGAATTTGCAATCGAAACGAGAAACTTCATCTCAGTCG
 GATTAGCCCATACGTACGTATGTTGGTTGTACTGTTGAATCCGGCTCCAAAATGGATC
 GCAAAGAAGCTGTTTTGAAGGACCAAGCGGTACGTGAAGAGCTCTCTCCAGCAAGTGCA
 ACCAAAGGCAAGAAGCACAAAGAAAGTCCTTGCTCTCCGGGAGACACCAATGCTGGCCA
 AGTGATATCGTCGGTGGATGAAGACACCAATGCTGGCCAAGTGATATCATCGGATCAAG
 ACGCCAATACTAATACCAAGAACAAGAAACAGAAGAAGTGCCATTTTGCAAAGGATAAC
 AATGGCTATCAAGCTGCTGGTCTCAAGACAACCAAGGTTCAATTTGGCAATAAACAGGG
 AAGTCTCAACTCTACGGCTATAACCTCCAATTCTTCAATGACGAGTAATGGTCTGCGAGG
 GCATGCAACCAAGCTCAGATGCTCCCGCAATCTCCAGTATATCCAGTCCGAGGAACCAT
 GCCGCACATGCCTCTGTTAGCAATAAATCGGAAGAGACAGCTGGTGGATCTCAATCGCG
 GGTGCAATCGAATGTTGCATCCATATCCTCGCCACACCAGCAGAACCGCGCACAGCGGAG
 TCACCTGCGCCGAATCGCAAGTGATACTAAGCAGGGACGTTAATGAAGATGCCGTGG
 TGCTGGTGAGCAGCGGCGATGAGGAGACCACCGCCGCGACGATGGCCAAACAGAGAGA
 AGACTCACCCGACGAGAATCAAACCTTTTTCACATTTTCGCCAACCGGTACCGGTGGCC
 TGGCATATGCTCTTCTTTAAATTTCTCCTCCTCGCTATCCTCATCCATCACTAGTTGGC
 CATTTAATGAGGTAAACTTAATGACCGGCAAAATTTTGCCTGCTGATTGCTTTCATTC
 ATTAGCTTCTGGGTGAGATTAGCGATATCCTTACGGTCTTAGTTACTGTGAGCATGTC
 AAACCAATTGCTCAGCTCCCAATGGGCAGTGATATAGTTGATGAAGGTTTTGTAAAGA
 ATTGCTCCACATACTGCAGCAGATAGAGGCCGCAATCTGTGAGATTCTGCTGTTGCGGC
 ACCTCATAGAGTAGAGAGGGCATATTATCCTTGTGAAGATATGCGCCAGCGCATTCGC
 ATACTTTGCCCTTGATACAGGTGAGATAATCCCGCAATATGGCAATCGCTCGATGTCTG
 TCAGGTGACCGCCAACGAATCGAAAATAAGAATAAATGGCTG

***diver2* (delecionado)**

TGTTTAGCTATTGAGTTTTTTTAAAAGCTTTTGTGAATAATAAATAGCCGAGGAAAAGTTT
 GTTTATTTTACTTTACTGTTTATTTTCGCAATTGTTTACAGCAGCTAGGCCCGACAAAG
 AGCTATATCGAATGCACAAGTTTAGTCTCTTCCGACAGACGAAGAAGTGGAATTTGGCG
 GGACAGACAGCCGAAATGCAGCGCCCAAGCTTAACGTAGGTAGATAACAAAAAGAACAA
 GGAATGAGCGCCGTACAGATAAATGCGTGCGGAGAACAGCGGTTTTGCCCTTTGGGCATCG
 CAACTGCTACCACTGACTACTTCGGCTTACAATATTA AAAAGTATGAGATTAGTCTACT
 GCACAGTTTTGGCGGCTGCATGACCCAACATCTTCTCCCGTTGGAAGCTTGGATTCCA
 ACAGAGTCCTCAATGGGAAGCAGACACAGCTTGTTCACTGCCCCGCTTATTACTCCAGA
 CGCGGTCTCAATTCTGCAACTCGAGCGACGCCGTCTCTGCCAGGAATCTACTGCATGT
 CCTCGCCAAAGGCCATCTCATTGGGGGTAGATTCTCGTCCTTAACAAGAACGACGTTGT
 CCACGGCTACGCCAGGCTTTGGGGTGCGCCACTTGGAGCGCTGCTGGAGCAACGTCAAG
 TACTCTTCCTTCATCGCGACCAAAATATTTGCTGAAGAAAGGAGATGCGTTGCCAAGA
 GTCAAGCCGATTATAGTTAAGGCCCGTTATATCTGGCTCGTCAAACGACGAAGCGGAC
 CACCATTGAGAAAATGCGCCGCTGAGAAATCCGAAGTGACAGACCGAATCGCAATTCTCT

ACCGAGAATCCATCAGTCGCGAAATTGGAATCCCCGATTACAGTGACGGAGCCGGACGA
CCTCGACCTCTTAAGTTGCAGTTGATTTGCAAGCCGAGAGGTGACCAAGTGCAGTTGAG
AGCCAGAATCTAACAAGGCCCTGCAGGGAACGAACAGTCCTGACCGATTCTGCACTAGA
ACGGTTGCAGTGGCTA

CG12717, -2L35, -2L38

TCCAAAAATCATCTGTCCGAGTTGCTATCGCCATCGGAGTTCTGCTGCTCATATCGATA
CATATATGTTTCCAATTGATATGGATTTTGTGGGTATTTTTTAAAAATATGTATCGATA
GTCCGCTTTGAAATTTGAATGTTGGGAATATTAAAAATAAATAATTTGAACTGACCA
TTTGATCATTTACAAAATTAGTATTCTTACATATGGTATTAAAGAAAGTTAAAAAGTTCAG
CCGTTTGTGTGCAAGTTCTTTGTCTGGAAATTTCAAAAACAAAAATGTATTCTCGTATTA
ATTTTCAGCATTTTGCAAAATACAAGTACAGACATTAATCAAATTATAGTTGTGCGTCCGT
TATCGGGTCCACTTGTCCGATTTTTAGTCGTCAGCGTTGTATTGTATTGAAACACATTT
GTTCAATTCTTTGCTAATTTCCCGTAAATGTGCAAACGACAGTTACCGCCAGTTACCGT
AGAAAGAAAAATCTCGAATTAATACCGAGGTTGTCTATTTGTGAAACTGGGGTGTAGAC
ACCGATCACTTTTAAATCAACTTTTAAATCAATTTTGAATGCCCTAATTACTTCTCCTTG
GTTTTCTCTGAAATCCTCACCAAATATTAACTTTGGTTTAAAGAAATAATCAGTCAAAA
TAATATTTTTCGTGAGAACCGTAATACAATTTCAATTCGGTGATTCAGAAAACTGATA
ATAGTAGGGGAGCCGGGTGAACCCCGCCCGTTACTTTGAACGACATAAAGAACGTGGT
TCACGCCGAATTTGGGATATTGAACTTCGTTTTGCTTAATAAAGACGAAGATCTGCTCC
AGTACCGTGGCCTGATGACACGTTGCTGTAATTTTGTACAGAGGTGAGGGAATTCCTCC
CTGGATAACGGTGTGCAGGAGCTGGTAGCGCTAATGTGCCCTCCCCGACAGCAACAAA
TTGAAAAACGACGACCACACGTACTAGAGGGATGCATAACACTATTCTCCAGAAGCGAT
TCGTTTTACTGCGTATTGAGAATACATGCCATTCAGTTGGTCGCGTTCTCTTTGCGTTC
ACTTCGTATGTATTCGGACTTACAGGCGTCCGATCGAAATAAACTAACTGAATGTGTTT
TAAGAATGAATGAAGCGAATGAATTTTCAATGGTAATTCAGAGTACTCTGAATCTTCA
TGTTGTAATTGGGTAAAATGAGTTTCGGACAGCGCGGAGGTTAGTCAAAGTTTGTGTTTT
ATTATGTTTATTGTTGTTATTTATTTGCGTGTATGTCTTATTTAAATAATTTGTAGCAT
TGTTTAATATAAATTTATTTAGCGATCTTGTCAAATTTGTGTTGTCAATTATAAATATTTT
AAATATAATAGTAATAAAAAATAATAGAATCGTTTGTATAGCTAAAGTCTGGCGACGCAA
GTTAGAAGGTTTAAACACGTAGTTTATACAGGATCTCGTAGAGTGAAAGTATATACTAC
ACTCTTTGAAAAGTATGTAACAGGTAGTTTCCGACCATTTAAAGTATATCTATGCTCTA
CCGAGTTTGTGTAGCCGAAAAGAGTTTGGCACGAGCAATAGTGGGAATAAAAAATAGTTTC
ATATAATTTATGCCACCTATGTAGAATTTACGACGCAACCTAACAAAAATGATGTTGCCC
AGGATGCGCATCTCTCCACGGTACATTTTTTATACTAATATTATAATATTATCTTTTAAAG
CTACGCACAAAAATATTACAAATCAAATTTCTTCTCAGTTCAGATTTTATAAAAACTAA
ATTCAGAGGGTAAGTGAAAAGTGCAATTTGTGTGGACCAAAATTTCAATTTATTTTAGAC
CAACCGCATAAATTTTCCATTGGTGTGCCTTGTCTGCCAATCAGTTTAAACTTTTTC
CATCACCTTACGTTTAAAGATTCTCTGGTTAAGAAGCTTTTTTCCCTGAATTTCCAGTTCT
GTTAAATTCATTGTGAGTCGAAGCGGAAAATCGGCGTCCATCTGGCTTTAACTAATG
CTGGGGGAGTCCATTGACAGCCAACCTCCTGGCCACCCCATACCTGTAATGACGAC
ATTCGCAACAACACTGGCAGCAACGCGGGCACTTAAATAAATCATAGTTTCACGCGGTT
TGGCCGAATGAAATATTGAGTTATTCTAAAGAGGAACCTCCTGGTGATGAAGATTTCAAA
TTTATATAGTATTTTAAATTAATGTTTGTCAAGCCCAAAATATCAGCTTGGTGCCATAT
TAACTCGTGAGGTCCAACCGGAAAAAGCATTGAGACGTAAACCCACAAGAGTTGATATC
AAAGGCGACAAAAGCGTGGAGAATGCTGAAGACTTAAAGCGATTTCAGTCAGATCGAAAA
GAATGCCATAAGAATCTGGAACAAACACAGTGAATTACGAAGCGATCTGTTATAGGAGG
AGGAGCACGAATAGATTACCTTTAGTCGGTATCATCATCCTTATAATTTGTTGTTGTT
TGTTTCGAGTTTTCCGACCCGTTACTCAGCTAGTGTAACGAACGCGAAATTTTATTA
CTTTTCTCTGATATCGATATATATTAAGAAATAACATGGAAAAATTTAAAAATTTGTC
CAAAGCGTGGGTGTGATTAGTTTGCCGACTTTAGAGCGTTAAATGGGCGTGGCAAAAA
GTTTTTTGACAAATCGTTAGAAATTTATAAAAGCCGACGAGAAGTATTGAATATTTGAA
AATAATAATAGATAATAATAATAATAATAGAGTGCCATTTCTTTTCTACTCACTTTGGA
CACCAAAAAGTTATTAATAATAAAGGTTGGCCATCCCAATTTTTTTTATATATTCGG
TATATAAGATAGTTATTTTGTAAATAAATAGTGCAGATATATATATAATT

-Idefix

AAATACTTGCAATATTTTATAAGTGAAAAAATATAATTTTTATTTCACTAAATGATTA
 CAATCTTTTTATTTGAGAAGAACACCGACCTTTTACAATTAGGTCTCAAACCTGAACTCT
 AATAATTATTCTGATTTGATTGACGTTCAAGCCTCGGTATCGAGCTTTTCTAATATGGA
 CTTTAGACTTTAGGATTCTGATCGTGAAGAGAAAAAGCTTTGCGTTTATCTTGGATTCA
 AGTGCATTTGCTCTTGGATACAAGTGCTCTTAGATTCTTATAAATTTGTTTATTGAAATC
 GATACTTTTGTTTTTCGGGATTGCGGCTCCAAGCTTTGAACGTGGGAATGTGACGATGG
 GGTGAGGCTTAAGCAAGGAATTGTTTAAATTAGGGGACGGATGTTTAGTCTACCAGTGG
 GTGACTTATCTAGAATTGGGGTTTTTTTTTAAACAATCTTATTCTTTCTGGTATCTGTTGG
 GTACATCCGGAGTAGTGTATGTTGTATTGGGGTGTATATGTTGTAAGTGATTTAGTATG
 TAGGCATATGCTTAAAGTCTTAGGGACTTATGGATATGTCACCTCCCCAACCTTTGAATT
 TGGCCTGTCTCAGGTCGTTAAATTTGGGTATAAGGTAATATTATTTCTTGTGTAATA
 TTTTCATTGACATTATTTTCAAGGTGCGTGTGTTCTTTTTGCTTACGATATTTGACAAT
 AATTTTCTTAGGTATATTTTTTAACTTATATGCTACATACAGAGTCAAACCTAATTATCA
 TTATGACGAACGTTGTAATGCATATAATTTGGAAAAAGTTATAGTGTTTTACATGGGAT
 TCTATCATTTCTCTCGTTTGTAAAGTGATTTATTGGTTCTAGACGTGTTACGTTGTTGGT
 GATATAAATTTGTTTGTGTAAAGTCTGCTAGATTATTAGAGATTAAGAATTCATCTATTT
 GCATGGTACAATTTTTAAATTTTAAATGATGTTGTTTCCTGAAATTAATATTTATTGTTT
 GTACAGTCTTGGTTAACAGCTGTTTCAGGAATATTTCCATGTTAATAGTATATTTGGTTC
 TATATAATTTATTTTGTAAATTTTTATGTGTTTTAATATATTTGCATTGAGTTTGCTCTT
 GTTTGATGATTCCAATAATACATTTGTGCAATATTACCCATTGGTATCTTTATTGAAT
 ACTTGGTTATCAAATATGTAGAATTTATCGAATATTTGCTCGGTTAGAATATAATTATG
 TTCATCTGGGTACGGAACAATTTGGAATATTGGAATTTTAGTGACTTCGCTACATGGGA
 AATAATCAAAATTTTCGTTTCGTATCTGTTTAAAGCCAAGTTGACGTTTTTATCTTTAGCA
 TTTTTTCTGAATTTACATGTTTTTAAATAGTCATGCTTTAGTAACTTTGGATTGAAAAAT
 CCTAGTCTGGTTAATTGCATACTCAATTCATGTCTTCTATATATTTCTGTAAATTTGCTC
 TAGGTTAAATATTAGTACCGCAATTTGTTGGTGTCTGTTCTTTATCTACTTTAAGCCTTAT
 TAATTATATCTATACCACATTTGATTACATCTAGAATCGTGTTTAGGTATCTGGGTTTTT
 ACAGAGTCTTCTGACATGTTGTTTATTTTTTCTTCTAACCTCTTCTCTGTCATCCTCATC
 TAATGTGCCGAATAAGTATTTGTATGCTTTTCTACTACGTTTAAAAAGACCTCTTTTAC
 TTCGACTAATAATTCTTAACCCATTTATTTCTCGTTTTAATTTGTCGACTAAATATTGT
 ATTTGTGGTACATTAGGATAGTTGTTTGCTTCACTTACTATATCTTCGAACATGAGCAT
 TGTCCTTTGTTATGTTTATGCTTAAATAATGGTATTCATAGCTGGTTGGAATTTCCATTG
 TTCTCTGTTTGGAAATATAAGGTATCCATTTTTTCGCGTTTATTGGATTTACTTCTATATTG
 TTGCACTTGACCGTGATGATAAGTGATAGTATGCAACTAAGCATTATTAGAAGTATGTT
 GTGTACTTTGTGCTTCTTTGGTTGGTCTGGAATTATCATATTTGCTCTTATTAAACCTT
 TTCTGTTTCTTAAATTTTGACTTATAGTGAACGATTTTTCCTGCCACGATTTGTTTCTTC
 GAAATGTTTATCATCTACTTGTTCATTCCTTCTGCTTTTTTAAATGGATTTGGTTCAATT
 TACATTTTACAAGTGGGGCTTGCTTATATTTTATATCAACTTCAAAATCGTGTCTATTC
 TTATTGAGGTATTCTATCCTATCGATTTTGTGTTTGTGTACATTAAAAATCTGGACTGCC
 TGCATATAGGAAAATGTCTGCTGGAATCTTTTAGTACTATTATGATTAGTTTTGTGAT
 TGATATGTATAGAATTGTTTCGAATTTTGTCTTTCTTCAATATTTTGTTCGCTACCAA
 TGATTCTTAGCTTTTTCGTTTACGGTCTTGTGGAATCTTTCTATATCAGATATACCATTT
 TTGCTAGTCGTTATAGAAATTTGTACACCTTCAGATCTTAACCAATTTTGTAAAGGCTAT
 ACACATAAAAGCTGAGTCTTTGTCTGCTTTAATTTCTTGGGTTTTTCCATATCATTGA
 ATATTTTAGTAAATGGCTCTTTTGTCTTCTAGCCAATCTCTACTTTTACTTCAACTAGT
 GATGCAAAATTTTGAATAGATATCAATGCAAGATAAGAAGATCTGGTTTCTCTGTGAGATA
 AAAATCTATCACGTATTTTTCTCTTGTGTTAAATATTTCTGGTGTAGTCTCATATGTCA
 ATTTCTGTGTTTCGATGTTCTGTTTTTGTCTAGATTACAAATTTTACATTCATTGATAATG
 TTTTGAATAAGCTTGTGACTATCAGGATAGTAATATTCTTCTTTAAATAAATTGATTGT
 TTTTCTATACCCGGATGTAAAAGTTCTTATGTTTCTTTAAGATTAAATCTTTGAATT
 CTGCATACGTTAGTATATCAATTAACCTTTGTGGTGCATCTTAAAGTTTTGTGAAATTG
 TTAGGGCTTATAATTTTCGGTAAATGCCCTCTGGAAGATCAGAAAACTTCGTCATTAGG
 TGAATAAATTTGCACTCTTTTTGGTACACACATATTCCTTAATGATGTTTTTGGCGAGTT
 CAGGTGTCATTTCTTTTAGATTATTTTAACTTTTAACTTTTGGGAAGTAATGTTGTACA
 CATGTTGTGTCGTTGTGCGCTTTTTCTCCATTCGTCCTAGAGAAGAAATTTATTGGTCTT

TCTGTTATAGATATGTAATTTTGATTATCTTCACTGGCGCTGTGTATAGTTGCGTCTGC
 GCTGTTTGCGACCTCGCCAACCATAAATTCTTCTATTTTCGTGCGGAAAAGAGCATCCG
 CAACATGGTTTTCTTTGCTTGAAAGATATTTGGTTTTGTAATCAAATTCATTGAGTTGT
 ATTTTTCATCTTTGCAATTTTCATGTTTGTTCTTTAATGTTGTTGAGCCATACTAGTGG
 CTGTGATCACTTAATATTTCAAATGGCCTGCCGAATAGATATGACCTAAAAATATTTTG
 TAGCCCAAACATAGCTAACAATTCTTTTCAATCGTAGCATAGTTGATTTTCATGTTTCG
 TTTAGCGTTCTACTGGCATAACAAACTGGCTTGTGATTCTGTGATAACACTGCACCGAT
 AGCTACATTGCTGGCATCAGTTGTGTCAGAGAAAAAGGTTTTGAAAAATCAGGATAGATTA
 ATATCGGGTCTGAAGTTATCAAAACTTTTAATTTTCAAATGATTCGATGTAGTCTTTA
 CATTTGATGTCTATTACAGCACCTTTCTATAATTTGAGGGTCATGGGTTTAACTATTTT
 GGCAAAGTTAGGAATAAACTTGCGATAGAATCCACATAGTCCCAAAAATGATTTTATTT
 GCTTAGGTGTCTTGGGTAATGGAATTTGTAATTGCTTTGGTTTTATTGGAATTTGGT
 TTGATGCCATTTGTTGTGACGATGTGTCCTAGGAATTCAGATTCTTCTCCTGAATTC
 ACATTTGTCTAGCTGCAATTTCAAATTAGCGTCTCTCAGTTTTTCAAAGTCTTTCTTTA
 GGGATAACATGTGTCTTCCAATGAAGTGGAATAAATAAATATCGTCTAAATAGACT
 AAACAGTCTTTGTAGATTAAATCTTCCAGAAGATTATTCATGCATCTCTGAAAAGTAGC
 TGGAGCGTTTTTTAAACCAAAAGGCATACGAGTATATTCATAATGCCCATGCTTAGTTG
 AAAAAGCTGTTTTTGCAGTAGAATTTTCATCCATTTGGATTTGGTGGCATCTACCTAAT
 TTGTCCAATATCTCATCCATTCGGGAATGGGAAATTTGTCGTTAACAGTTATCTCATT
 TAGATTCTGTAAATCGACTACCAACCTGAATTTCTGTTTCCCAGAGGCATCTTCTTCT
 TGGGGACCACCAAAATAAGAGAACAATCCCTTGTTCATCATTCTTTAATTTGTTTGT
 TGACTTCTTGGTCAACGCTTTGGGGGTACTTGTATGGTTTACGGTATACTGGGTCTTCG
 TGTTGAGTTTGGATGACATGTTTAATAGTACTGGTGAAGGTCAAATTTTCGCCCTCTTT
 GTACTGAATGTCTCTATATTCGTATAGGACCTTCTTTAAACATTCAACTTCCTCTACAT
 TTTAGTGTTCGAGTCTATACTCGTTACATTCGCGTAACTCATTGTTAATCGCGAAGTTG
 ACTATATCGTTGTCAACGGACATGGGATCGAGATGTGTTACATAATTGTCCACTGTGTC
 TTTAACAACATTTGAAATGAGGCATTTGAGTAATGCAGTCTCCTTCTTCTGCTGTTGAT
 GCTTTGGTTTAAAGGCATTAGGTGCGGTCTTAACCTTTTGCATTTTGGTTTCTTTTCAT
 TCTTTTGTGTGGGTCAAGGCATTTGGATGCGGTCTCGCCCTTCTTTCTTCATAGAGA
 AACTTAAA

